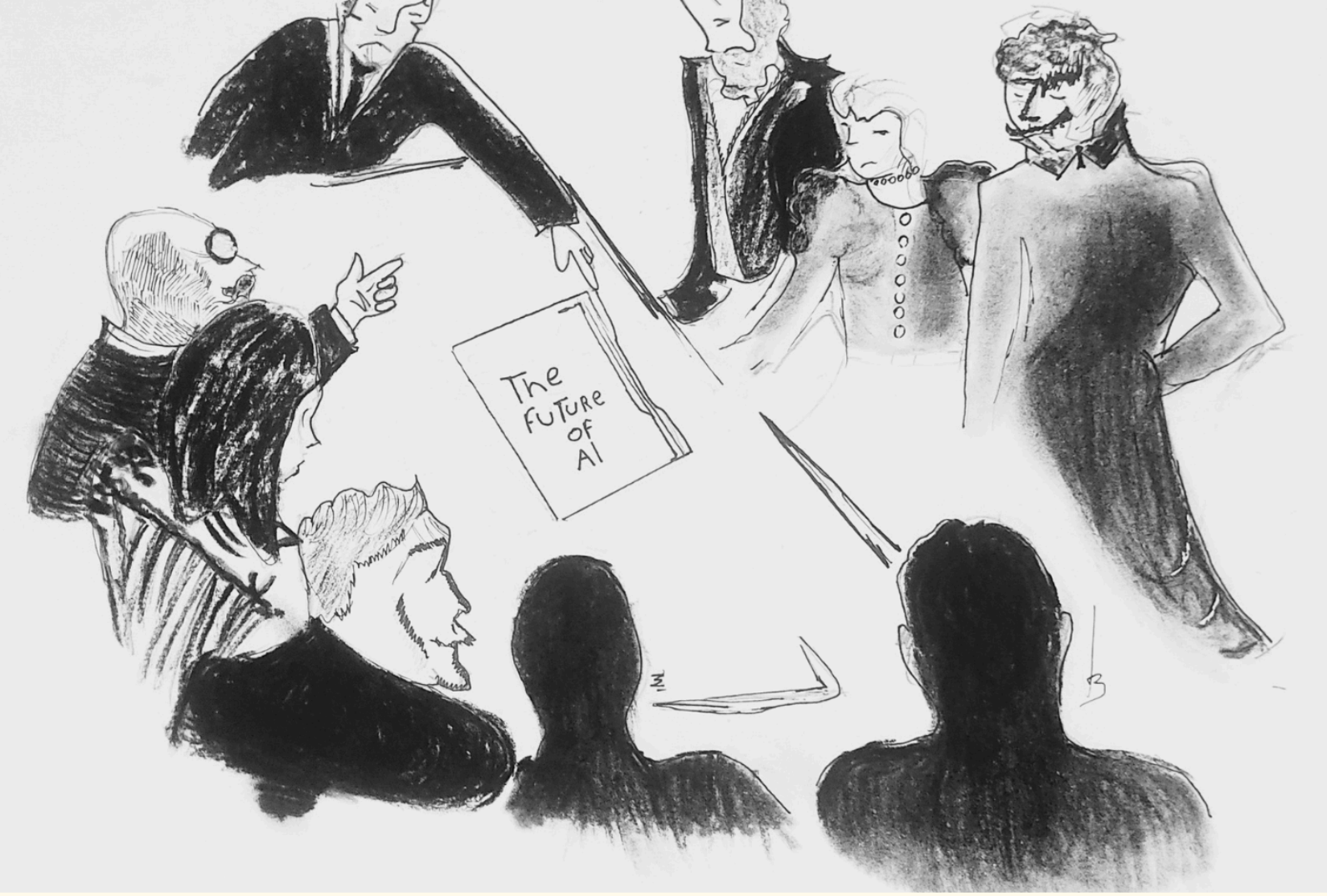


UNIVERSIDADE FEDERAL DE GOIÁS/INSTITUTO DE INFORMÁTICA

André Martins Dantas

Impactos Sociais da IA

Desenvolvimento e Validação de Framework Analítico Multi-Dimensional

UFG
UNIVERSIDADE FEDERAL DE GOIÁS

UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
INSTITUTO DE INFORMÁTICA (INF)

ANDRÉ MARTINS DANTAS

Impactos Sociais da IA

Desenvolvimento e Validação de Framework Analítico Multi-Dimensional

Goiânia
2025

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TRABALHO DE CONCLUSÃO DE CURSO DE GRADUAÇÃO NO REPOSITÓRIO INSTITUCIONAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio do Repositório Institucional (RI/UFG), regulamentado pela Resolução CEPEC no 1240/2014, sem ressarcimento dos direitos autorais, de acordo com a Lei no 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo dos Trabalhos de Conclusão dos Cursos de Graduação disponibilizado no RI/UFG é de responsabilidade exclusiva dos autores. Ao encaminhar(em) o produto final, o(s) autor(a)(es)(as) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do Trabalho de Conclusão de Curso de Graduação (TCCG)

Nome(s) completo(s) do(a)(s) autor(a)(es)(as): ANDRÉ MARTINS DANTAS

Título do trabalho: Impactos Sociais da IA

Desenvolvimento e Validação de Framework Analítico Multi-Dimensional

2. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador) Concorda com a liberação total do documento [X] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante: a) consulta ao(à)(s) autor(a)(es)(as) e ao(à) orientador(a); b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo do TCCG. O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;

- Submissão de artigo em revista científica;

- Publicação como capítulo de livro.

Obs.: Este termo deve ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **André Martins Dantas**, **Discente**, em 16/03/2026, às 19:36, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Fernando Marques Federson**, **Professor do Magistério Superior**, em 21/03/2026, às 09:33, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

ANDRÉ MARTINS DANTAS

Impactos Sociais da IA

Desenvolvimento e Validação de Framework Analítico Multi-Dimensional

Relatório final de Trabalho de Conclusão de Curso, apresentado à Universidade Federal de Goiás, como parte das exigências para a obtenção do título de Bacharel em Inteligência Artificial.

Orientador: Prof. Dr. Fernando Marques Federson

Goiânia

2025

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

DANTAS, ANDRÉ MARTINS
Impactos Sociais da IA [manuscrito]: Desenvolvimento e Validação de Framework Analítico Multi-Dimensional / ANDRÉ MARTINS DANTAS. - 2025.
124 f.: 2025

Orientador: Prof. Dr. Fernando Marques Federson
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Goiás, Instituto de Informática (INF), Inteligência Artificial, Goiânia, 2025.

1. Inteligência Artificial. 2. Impacto Social. 3. Framework.

I. Federson, Fernando Marques , orient. II. Título.

CDU 004

ANDRÉ MARTINS DANTAS

Impactos Sociais da IA

Desenvolvimento e Validação de Framework Analítico Multi-Dimensional

Relatório final de Trabalho de Conclusão de Curso, apresentado à Universidade Federal de Goiás, como parte das exigências para a obtenção do título de Bacharel em Inteligência Artificial.

Data da Aprovação: 09 de dezembro de 2025.

Prof. Dr. Fernando Marques Federson
Orientador (INF-UFG)

Prof. Dr. Aldo André Díaz Salazar
Coordenador de TCC do BIA (INF-UFG)

Prof. Dr. Anderson da Silva Soares
Coordenador do BIA (INF-UFG)

Prof. Me. Lucas Araújo Pereira
(INF-UFG)

ANDRÉ MARTINS DANTAS

Impactos Sociais da IA

Desenvolvimento e Validação de Framework Analítico Multi-Dimensional

RESUMO

Este Relatório de Conclusão de Curso tem como objetivo reunir os resultados da minha jornada para me tornar um especialista em **Impactos Sociais da IA**. Uma ilustração e sua narrativa descrevem os períodos de trabalho. Os Apêndices contêm os Termos de Aceite de Entrega e os resultados obtidos durante cada período de trabalho.

Palavras-chave: Inteligência artificial; Impacto social; Framework.

ABSTRACT

This Course Completion Report aims to bring together the results of my journey to become an expert in **Social Impacts of AI**. An illustration and its narrative describe the work periods. The Appendices contain the Delivery Acceptance Terms and the results obtained during each work period.

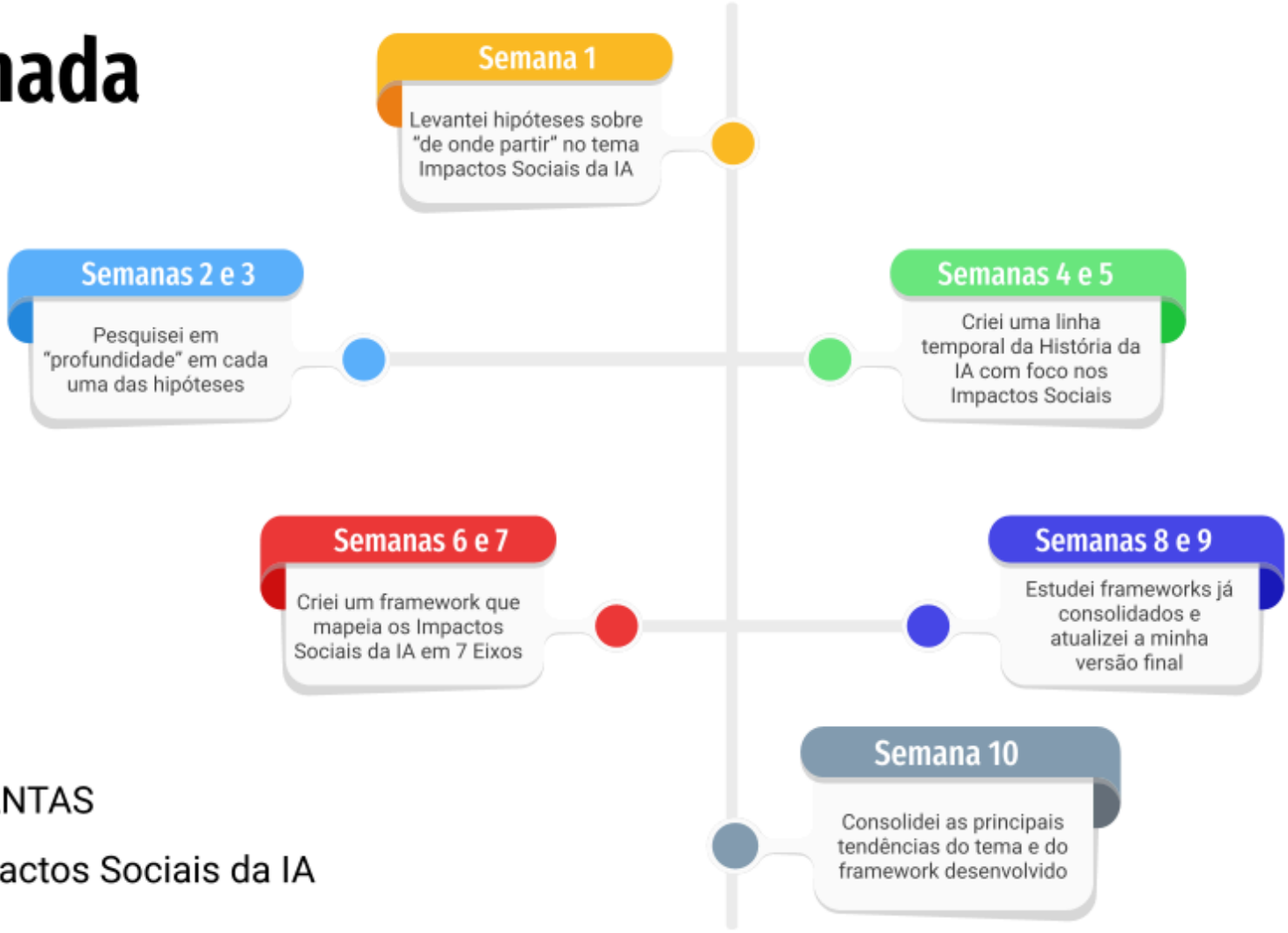
Keywords: Artificial intelligence; Social impact; Framework.

Goiânia

2025

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

Minha Jornada



ANDRÉ MARTINS DANTAS

Especialista em: Impactos Sociais da IA

Instituto de Informática (INF) - UFG
Alameda Palmeiras, Quadra D, Câmpus Samambaia
Fone: (62) 3521-1181 / Fax: (62) 3521-1182

MINHA JORNADA

Nome: ANDRÉ MARTINS DANTAS

Especialidade: Impactos Sociais da IA

Objetivo deste documento

Durante o processo da disciplina Residência em IA[1], foram gerados diversos resultados na construção da minha especialização. A cada semana, um conjunto de resultados foi formalizado por um Termo de Aceite de Entrega e avaliado por uma banca, considerando o planejado e o realizado para o período. Este documento tem como objetivo descrever esses resultados obtidos, fazendo referência aos Termos de Aceite de Entrega e seus documentos associados.

Minha Jornada

Minha jornada iniciou-se na **Semana 1** com um esforço de delimitação ontológica para compreender o que realmente significa o tema "Impactos Sociais da IA". Parti da desconstrução do termo, estabelecendo três premissas fundamentais para o estudo: que existe uma "IA" distinguível de outras tecnologias, que ela possui capacidade de agência para causar impactos e que tais impactos transcendem o técnico, sendo fundamentalmente sociais. Para estruturar o campo de conhecimento, levantei três hipóteses temporais sobre a emergência do tema: a Hipótese 1, de cunho sociológico, sugere que o tema sempre existiu como um processo contínuo de avaliação de tecnologia; a Hipótese 2 situa o nascimento do tema junto com a criação do termo "Inteligência Artificial" na década de 1950; e a Hipótese 3 argumenta que o campo é um fenômeno recente, derivado da onipresença do *Big Data*. A decisão metodológica de utilizar a Hipótese 2 como ponto de partida para evitar tanto a abstração excessiva quanto o presentismo, juntamente com o detalhamento das premissas ontológicas, está documentada no **Apêndice 1**.

[1] Dez Semanas, entre setembro de 2025 e dezembro de 2025.

Nas **Semanas 2 e 3**, dediquei-me à pesquisa em profundidade, conectando as raízes da Cibernética de Norbert Wiener — que já via o cérebro como modelo de computação — aos desafios contemporâneos. Analisei como a questão teórica de Alan Turing ("Podem as máquinas pensar?") transformou-se, com o experimento ELIZA de Joseph Weizenbaum em 1966, em uma crise prática de discernimento humano, onde o perigo não era a inteligência da máquina, mas a ilusão de compreensão projetada pelo usuário. Avançando para o cenário pós-*Big Data*, o estudo identificou limitações críticas na IA atual, classificada como "IA fraca", que sofre de problemas como a falta de "ancoragem de símbolos" (o computador não conecta o símbolo ao significado real) e a ausência de uma verdadeira "Inteligência Social Artificial" (ASI). A pesquisa também mapeou como plataformas como Facebook e LinkedIn deixaram de ser apenas canais de interação para se tornarem ecossistemas moldados por algoritmos que inferem comportamento social, gerando riscos de vieses e violações de privacidade. Todo o referencial teórico, incluindo a análise do "problema do *frame*" e a transição histórica dos impactos, compõem o **Apêndice 2**.

As atividades das **Semanas 4 e 5** concentraram-se na criação de uma linha temporal analítica, visualizando a história da IA como um "funil" que converge da filosofia e dos autômatos anteriores a 1891 para os marcos da computação moderna. Identifiquei eventos cruciais, como as Conferências Macy (1946-1953), que estabeleceram a linguagem comum da Cibernética, e o Workshop de Dartmouth de 1956, onde o termo IA foi cunhado. Um resultado significativo desta etapa foi a identificação do "primeiro pânico social" documentado: uma reportagem do *The New York Times* de 1958 sobre o Perceptron, que já previa máquinas capazes de se reproduzir e o consequente desemprego tecnológico. A linha do tempo também mapeou os ciclos de "Invernos da IA" causados por limitações técnicas, como as apontadas no livro *Perceptrons* (1969), e a influência da cultura pop (como *HAL 9000* e *O Exterminador do Futuro*) na percepção pública. A análise culminou na convergência de dados (MapReduce), algoritmos (Backpropagation/LSTM) e hardware (CUDA) nos anos 2000, detalhada no **Apêndice 3**.

Diante da complexidade do cenário pós-2006, as **Semanas 6 e 7** foram dedicadas ao desenvolvimento de um *framework* próprio, uma vez que uma taxonomia simples se mostrou insuficiente para capturar a multiplicidade de atores, geografias e gerações tecnológicas.

Criei uma matriz de 7 Eixos Independentes, refinada em duas versões principais. A versão final incorporou uma "Era Fundacional" (W0: 1950-2005) e alinhou os domínios de aplicação com as categorias de risco do *EU AI Act*. O *framework* classifica impactos em quatro naturezas: Vida Material e Segurança, Poder e Instituições, Informação e Cultura, e Autonomia e Desenvolvimento Humano. Além disso, introduzi uma dimensão de "Profundidade" baseada no modelo SAMR para distinguir entre mera substituição tecnológica e redefinição de tarefas. Apliquei essa lente analítica para examinar quatro esferas da sociedade (Individual, Mercado, Estatal e Cívica), revelando vulnerabilidades como o "deskilling" de trabalhadores e a crise de capacidade institucional frente à IA. A estrutura completa do *framework* e a análise das esferas estão no **Apêndice 4**.

Nas **Semanas 8 e 9**, busquei validar e atualizar o modelo desenvolvido através do estudo comparativo de *frameworks* consolidados, como o *Algorithmic Impact Assessment* do *Ada Lovelace Institute*, a Recomendação sobre a Ética da IA da UNESCO e o próprio *EU AI Act*. Paralelamente, conduzi uma extensa pesquisa empírica, mapeando mais de 600 notícias de impacto social relevante entre janeiro de 2022 e outubro de 2025. Deste levantamento, selecionei e analisei eventos críticos que validam os eixos do meu *framework*, como a explosão da IA generativa e as greves de Hollywood em 2023, o uso de *deepfakes* eleitorais em 2024 (ex: robocalls de Biden e eleições na Índia) e os primeiros grandes acordos judiciais de direitos autorais em 2025. A pesquisa também cobriu impactos tangíveis, como o uso de IA para alvos militares em conflitos e a expansão do reconhecimento facial em aeroportos. A compilação destes estudos de caso e a comparação dos *frameworks* globais encontram-se no **Apêndice 5**.

Finalmente, na **Semana 10**, consolidei as principais tendências e refinei a versão final do *framework* com base nas evidências coletadas. A análise final destacou o ceticismo corporativo emergente, exemplificado por um estudo do MIT mostrando que 95% das empresas ainda não obtiveram retorno sobre o investimento em IA generativa, contrastando com a "mão invisível" das *Big Techs* influenciando políticas públicas através de *think tanks*. Concluí que a área de Impactos Sociais da IA exige uma análise que considere a interdependência entre a infraestrutura técnica (eixos de geração e mecanismo) e as estruturas de poder (eixos de atores e geografia). As conclusões finais, juntamente com as

recomendações de intervenção priorizadas pelo *framework* desenvolvido para mitigar riscos como a erosão epistêmica e a concentração de poder, formam o conteúdo do **Apêndice 6**.

Em virtude de todas as experiências vivenciadas nesta Jornada, gostaria de registrar que o percurso se revelou imensamente gratificante, não obstante os desafios inerentes ao processo investigativo. Contudo, a análise aprofundada aponta para um horizonte que demanda cautela: a evolução desta tecnologia, a despeito de seu potencial transformador, parece inclinar-se a uma lógica que privilegia a concentração de recursos em poucas estruturas corporativas, operando, muitas vezes, em descompasso com o bem-estar coletivo. Nota-se, portanto, o risco tangível de que a inovação permaneça subordinada a imperativos de rentabilidade e controle, em vez de servir plenamente aos interesses humanos. Sobretudo, o legado mais valioso desta trajetória foi o meu próprio amadurecimento crítico; reconheço que o pesquisador de hoje vislumbra as dinâmicas de poder subjacentes às ferramentas técnicas com uma clareza muito superior àquele que iniciou o projeto. Essa nova consciência, longe de encerrar meu interesse, apenas ratifica o compromisso de prosseguir investigando as complexas implicações deste tema.

APÊNDICE 1

Termo de Aceite de Entrega

Objetivo deste documento

Este documento faz parte do Processo da disciplina Residência em IA e tem como objetivo formalizar o aceite da entrega considerando o planejado e o realizado para o período.

Data da Reunião ("Gate") de aprovação: 2 de set. de 2025

Participantes da Entrega [matriculados em Residência em IA]:

André Martins Dantas

Entrega: [descrever a ENTREGA - requisitos e produtos gerados: links para textos, códigos, vídeos etc.]

Em minha fase inicial de pesquisa, realizei as seguintes atividades:

01. **Defini o tema central da pesquisa:** A origem, a definição e a consolidação do campo de "Impactos Sociais da Inteligência Artificial".
02. **Propus minha abordagem de investigação:** Estruturei a análise em torno de três hipóteses principais (H1, H2, H3) para identificar o marco zero do tema.

    (H1- se sempre existiu; H2- se nasceu com o termo IA, ou H3- se é um fenômeno recente).

    H1. Levantei as bases teóricas sobre tecnologia e sociedade: Investiguei o campo da Avaliação de Tecnologia (TA) como um precursor dos debates atuais. (Rip - 1986)

    H2. Levantei as referências clássicas sobre a gênese conceitual da IA: Estudei essa temática sob a ótica de **Norbert Wiener** (Cibernética), **Alan Turing** (a questão da "máquina pensante") e **Joseph Weizenbaum** (o experimento ELIZA e a reação humana à simulação de inteligência).

    H3. Analisei a consolidação recente do campo: Investiguei como o advento do Big Data e do *deep learning* criaram as condições técnicas e materiais para um novo domínio de estudo, distinto das discussões anteriores. Descobri sub-temas como a escala de dados, a penetração ubíqua de algoritmos e, crucialmente, a opacidade inerente dos sistemas de IA modernos (o problema da "caixa-preta").

03. **Sintetizei as conclusões das três hipóteses:** Defini a Hipótese 2 como o ponto de partida metodológico ideal, por permitir a construção de uma narrativa histórica que conecta a gênese conceitual da IA (década de 1950) à sua manifestação concreta e socialmente impactante nos dias de hoje.

A pesquisa completa e o desenvolvimento das hipóteses estão detalhados no documento principal do trabalho e pode ser encontrado no link a seguir: Fluxo_semana_1

Planejamento: [descrever o que pretende fazer para realizar a próxima ENTREGA]

Para a próxima entrega, irei avançar a pesquisa a partir da Hipótese 2, mapeando a evolução histórica do campo desde a década de 1950. O foco principal será a identificação e análise de *frameworks* teóricos propostos ao longo do tempo, utilizando-os como uma "lente" para estruturar o entendimento de como os impactos sociais da IA foram conceituados em diferentes períodos.

Observação: [caso precise fazer alguma observação, de qualquer "natureza"]

ACEITE DA ENTREGA:

CEDRIC LUIZ DE CARVALHO: Go!

Fluxo_Semana_1.doc

Para começar, precisamos entender melhor o que seria o tema “Impactos Sociais da IA”. Afinal, esse nome por si só não parece muito autoexplicativo; sobre o que vamos abordar, ele ainda parece ser um tema maior que engloba muitos outros.

Ontologia

Quando dizemos "Impactos Sociais da IA", estamos assumindo três premissas:

1. Que existe algo chamado "IA" distinguível de outras tecnologias.
2. Que esse "algo" tem capacidade de causar impactos.
3. Que esses impactos são sociais (não apenas técnicos ou individuais).

Será em cima dessas três premissas que vamos continuar nossos estudos.

Quando nasce o tema como tema?

- Hipótese 1: O tema sempre existiu, apenas mudou de nome.
- Hipótese 2: O tema surge com a própria ideia de Inteligência Artificial (IA) (década de 1950).
- Hipótese 3: O tema emerge como campo distinto apenas recentemente (década de 2010).

Hipótese 1: O tema sempre existiu, apenas mudou de nome.

A primeira hipótese, de natureza fundamentalmente sociológica, postula que o debate sobre os "Impactos Sociais da IA" não é novo, mas sim a mais recente manifestação de um processo social contínuo de avaliação de tecnologias. Sob esta ótica, a Avaliação de Tecnologia (AT) como campo formal é apenas um fragmento de uma dinâmica maior. Arie Rip (1986) fundamenta essa visão ao argumentar que os estudos de AT devem ser vistos como “parte de um processo social de avaliação de tecnologia, muitas vezes conflituoso, permeado por ações estratégicas...” (RIP, 1986, p. 415, tradução nossa).[2]

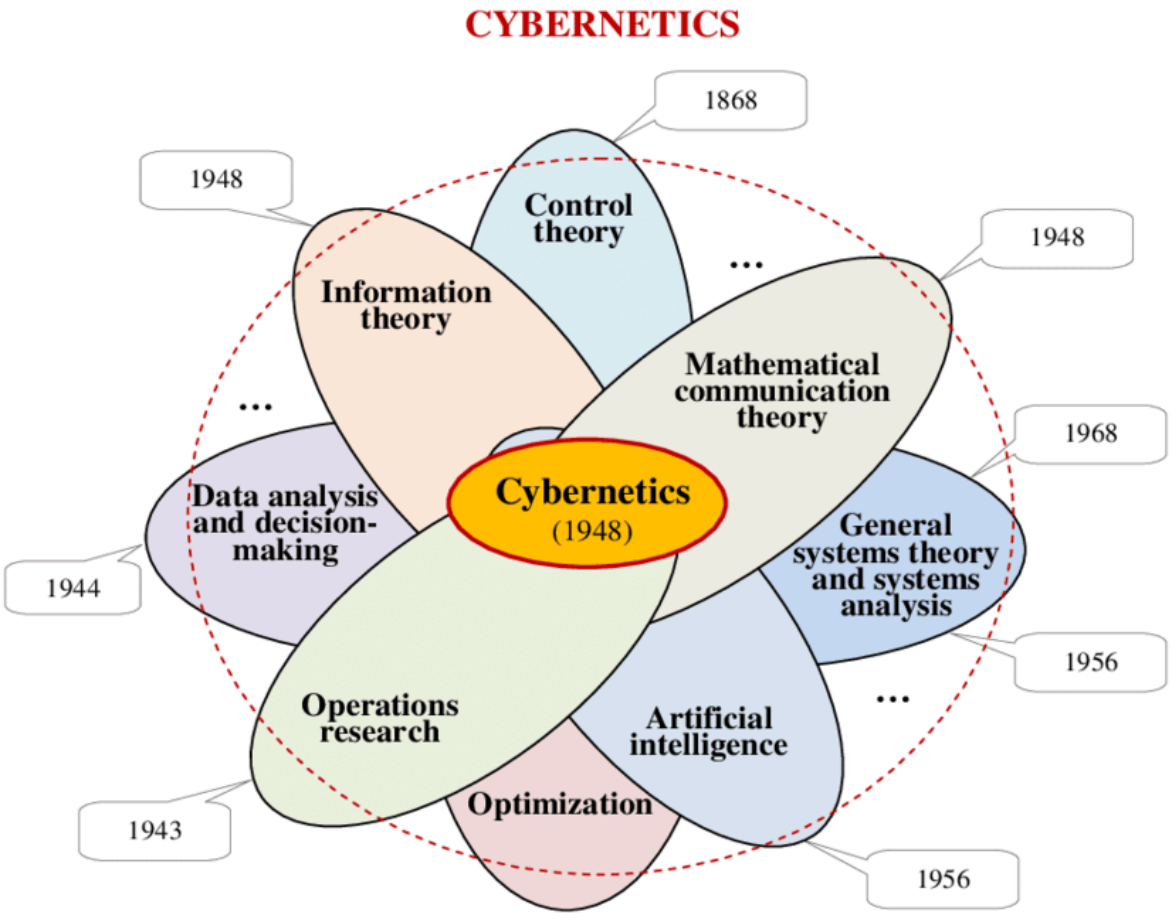
Nesse sentido, as controvérsias e debates públicos em torno de uma nova ferramenta, como a IA, funcionam como uma forma de "avaliação informal", onde “o aprendizado social pode ocorrer, e os estudos de AT podem contribuir para o processo de aprendizado” (RIP, 1986, p. 415, tradução nossa).[3]

---

[2] *“TA studies should be seen as a part of a societal process of technology assessment, often conflictual, shot through with strategical action...” (RIP, 1986, p. 415).*
[3] *“societal learning may occur, and TA studies may contribute to the learning process” (RIP, 1986, p. 415).*

Hipótese 2: O tema surge com a própria ideia de Inteligência Artificial (IA) (década de 1950).

O responsável por cunhar o termo Inteligência Artificial foi o cientista da computação John McCarthy. A primeira aparição documentada do termo "inteligência artificial" ocorreu na proposta "***A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence", elaborada por John McCarthy, Marvin L. Minsky, Nathaniel Rochester e Claude E. Shannon em 31 de agosto de 1955 (MCCARTHY et al., 1955).***



Antes de meados da década de 1950, pesquisadores em computação e neurociência já exploravam a ideia de máquinas pensantes, porém usavam termos como ***cybernetics*** popularizada por Norbert Wiener) e **teoria de autômatos** para descrever esse campo nascente.

fonte: https://www.researchgate.net/figure/The-composition-and-structure-of-cybernetics_fig14_287319297

Em seu livro ***Cybernetics or Control and Communication in the Animal and the Machine***, publicado pela primeira vez em 1948, Norbert Wiener lançou as bases para o campo multidisciplinar da cibernética. Quando se trata dos campos da cognição e da sociedade abordados no livro, estamos envolvendo o que posteriormente veio a ser chamado de Inteligência Artificial. Podemos destacar isto nos seguintes trechos declarados por Norbert Wiener em sua obra:

- **Cérebro como Modelo de Computador:** "...a máquina de computação ultrarrápida, dependendo como depende de dispositivos de comutação consecutivos, deve representar quase um modelo ideal dos problemas que surgem no sistema nervoso" (WIENER, 1948, p. 21, tradução nossa).[4]
- **A Natureza "Tudo ou Nada" dos Neurônios:** "O caráter 'tudo ou nada' da descarga dos neurônios é precisamente análogo à escolha única feita na determinação de um dígito na escala binária..." (WIENER, 1948, p. 21-22, tradução nossa).[5]
- **Neurônios como Relés:** "...os sistemas nervosos humano e animal... contêm elementos que são idealmente adequados para atuar como relés. Esses elementos são os chamados neurônios ou células nervosas" (WIENER, 1948, p. 165, tradução nossa).[6] Um relé é um dispositivo que pode estar em uma de duas condições, "ligado" e "desligado" (0 ou 1). A posição de um relé em um determinado momento é ditada pelas posições anteriores de outros relés no sistema, seguindo um conjunto de regras.

Ainda que os trabalhos de N. Wiener já explorassem o impacto que as máquinas poderiam exercer na sociedade, a questão foi abordada de forma mais incisiva por A. M. Turing em seu artigo '***Computing Machinery and Intelligence' (1950)***. Turing inaugura seu texto com a provocativa pergunta: "Proponho-me a considerar a questão: 'Podem as máquinas pensar?'" (TURING, 1950, p. 433, tradução nossa).[7] Contudo, ele rapidamente descarta a busca por definições literais de 'máquina' e 'pensar', argumentando que tal abordagem seria 'absurda'. Em vez disso, ele propõe que a questão seja substituída por um problema mais concreto e expresso em palavras relativamente inequívocas: um teste que ele denomina o

[4] *"...the ultra-rapid computing machine, depending as it does on consecutive switching devices, must represent almost an ideal model of the problems arising in the nervous system."(WIENER, 1948, p. 21).*
[5] "The all-or-none character of the discharge of the neurons is precisely analogous to the single choice made in determining a digit on the binary scale..."*(WIENER, 1948, p. 21-22).*
[6] *"It is a noteworthy fact that the human and animal nervous systems, which are known to be capable of the work of a computation system, contain elements which are ideally suited to act as relays. These elements are the so-called neurons or nervecells." (WIENER, 1948, p. 164-165).*
[7] Inicia-se seu artigo com a seguinte frase: *"I propose to consider the question, 'Can machines think?'" (A. M. Turing, 1950, Computing Machinery and Intelligence. Mind 49: 433-460.).*

'Jogo da Imitação', no qual um interrogador deve distinguir, apenas com base em respostas textuais, se está conversando com um humano ou com uma máquina.

Alguns anos depois temos a publicação do artigo (e famoso experimento) conhecido como ***"ELIZA, A Computer Program For the Study of Natural Language Communication Between Man And Machine"***, de Joseph Weizenbaum. A transição de Turing para Weizenbaum marca um momento crucial na história da IA: a passagem da questão teórica ("Podem as máquinas pensar?") para a questão prática e social ("Como nos relacionamos com máquinas que parecem pensar?").

Ao contrário do que se possa imaginar, Weizenbaum não queria provar que os computadores eram inteligentes. Sua intenção com ELIZA era, na verdade, satirizar e demonstrar a superficialidade da comunicação entre humanos e máquinas. Essa intenção fica evidente logo na introdução de seu artigo, onde ele afirma que seu objetivo é "desmistificar" programas de IA, explicando que, "uma vez que um programa específico é desmascarado, [...] sua mágica se desfaz; ele se revela como uma mera coleção de procedimentos, cada um bastante compreensível" (WEIZENBAUM, 1966, p. 36, tradução nossa).[8]

Para atingir esse objetivo, o programa simulava um psicoterapeuta, um cenário escolhido não por acaso, mas porque "a entrevista psiquiátrica é um dos poucos exemplos [...] na qual um dos participantes é livre para assumir a pose de não saber quase nada do mundo real" (WEIZENBAUM, 1966, p. 42, tradução nossa).[9]

A conclusão mais alarmante do experimento de Weizenbaum não reside no sucesso técnico de ELIZA, mas na reação humana a ele. O verdadeiro perigo, ou a "catástrofe" implícita, não é que as máquinas se tornem inteligentes, mas que elas nos convençam de que são, levando-nos a depositar uma confiança indevida em seus julgamentos. Weizenbaum adverte que decisões importantes são cada vez mais tomadas com base em dados de computador, e ELIZA prova como é simples "criar e manter a ilusão de compreensão e, portanto, talvez, de um julgamento que mereça credibilidade. Um certo perigo espreita aí" (WEIZENBAUM, 1966, p. 42, tradução nossa).[10]

---

[8] *"But once a particular program is unmasked, [...] its magic crumbles away; it stands revealed as a mere collection of procedures, each quite comprehensible." (WEIZENBAUM, 1966; ELIZA; Introduction)*

[9] *"...the psychiatric interview is one of the few examples of categorized dyadic natural language communication in which one of the participating pair is free to assume the pose of knowing almost nothing of the real world." (WEIZENBAUM, 1966; ELIZA; Discussion, p. 42).*

[10] *"ELIZA shows, if nothing else, how easy it is to create and maintain the illusion of understanding, hence perhaps of judgment deserving of credibility. A certain danger lurks there."(WEIZENBAUM, 1966; ELIZA; Discussion, p. 42-43).*

A catástrofe social prevista por Weizenbaum é, portanto, uma crise de discernimento humano (e portanto um impacto social). Ao nos relacionarmos com sistemas que simulam empatia e entendimento, corremos o risco de abdicar de nossa responsabilidade crítica, entregando decisões cruciais a algoritmos que, no fundo, são apenas "uma mera coleção de procedimentos" sem consciência, ética ou verdadeira compreensão do mundo real.

Hipótese 3: O tema emerge como campo distinto apenas recentemente

A terceira hipótese sustenta que o campo de "Impactos Sociais da IA" é um domínio de estudo relativamente novo. Sua emergência não é uma mera renomeação de debates antigos, mas o resultado de fatores técnicos e sociais sem precedentes, como a escala massiva de dados, a onipresença de algoritmos em decisões cotidianas, a opacidade dos sistemas modernos e uma velocidade de mudança que desafia a adaptação social.

Um fundamento histórico para esta hipótese é a revolução do "Big Data". Sistemas modernos, especialmente o *deep learning*, "dependem de conjuntos de dados massivos para treinar e melhorar os modelos de IA" (WINTER; DAVIDSON, 2019, p. 5, tradução nossa).[11] Esse novo paradigma material é evidenciado por um "aumento drástico [...] no número de pesquisas sobre Big Data e Inteligência Artificial" a partir de 2015 (BIRCAN; SALAH, 2022, p. 3, tradução nossa).[12]

Diferente de tecnologias anteriores, a IA contemporânea está profundamente integrada ao tecido social, mediando decisões críticas sobre crédito, emprego e segurança. Essa onipresença significa que seus impactos não são mais teóricos, mas diretos e cotidianos, pois os "algoritmos e os sistemas que eles suportam estão sempre inseridos em contextos sociais que os afetam e são afetados por eles em igual medida" (REA, [n.d.], p. 7, tradução nossa).[13]

Um conceito técnico fundamental para o campo é a **opacidade algorítmica**. Muitos sistemas de IA funcionam como "caixas-pretas" (*black boxes*), com "funcionamentos internos obscurecidos pela opacidade e complexidade dos algoritmos" (WINTER; DAVIDSON, 2019, p. 6, tradução nossa).[14] Essa característica não é acidental, mas muitas vezes inerente à tecnologia, pois a "característica inerente do ML avançado e da IA é a capacidade dos sistemas de aprendizado de manipular sua própria base algorítmica"

---

[11] *"Deep learning algorithms rely on massive data sets to train and improve AI models" (WINTER; DAVIDSON, 2019, p. 5).*
[12] *"as of 2015 a drastic rise is visible in the number of research on Big Data and Artificial Intelligence" (BIRCAN; SALAH, 2022, p. 3).*
[13] *"ML and AI technologies do not operate in a vacuum... algorithms and the systems that they support are always-already embedded in social contexts that affect and are affected by them in equal measure" (REA, [n.d.], p. 7).*
[14] *"Deep learning is a 'black box' (Pasquale, 2015), with its inner workings obscured by the opacity and complexity of algorithms" (WINTER; DAVIDSON, 2019, p. 6).*

(RICHMOND, 2020, p. 10, tradução nossa)[15], um processo que ocorre "além do limiar da percepção e do controle humano" (RICHMOND, 2020, p. 10, tradução nossa).[16]

Por fim, o campo define-se pela tensão entre a velocidade da inovação tecnológica e a lenta adaptação das estruturas sociais e legais. A IA está remodelando a sociedade "em um ritmo sem precedentes" (ZHANG; CHEN; HUANG, 2024, p. 1, tradução nossa).[17] Esse descompasso entre a rápida evolução técnica e a necessária deliberação sobre seus impactos consolida a urgência e a identidade deste novo domínio de estudo.

Conclusão e Direcionamento do Estudo

A análise das três hipóteses revela que elas não são mutuamente excludentes, mas representam diferentes níveis de abstração para abordar o tema. A primeira hipótese situa o debate em um plano conceitual e sociológico amplo, analisando a relação atemporal entre novas tecnologias e a sociedade. Em contraste, a segunda e a terceira hipóteses oferecem marcos temporais definidos. A Hipótese 2 ancora o estudo na gênese do conceito de "Inteligência Artificial" na década de 1950, focando em sua história e evolução teórica. Já a Hipótese 3 representa o ápice dessa trajetória, um ponto de chegada onde os impactos se tornam concretos e disseminados, impulsionados pela onipresença do Big Data e das interações diretas humano-máquina.

Para construir uma análise que não se encerre prematuramente nas aplicações atuais (Hipótese 3), nem se torne excessivamente abstrata (Hipótese 1), este estudo adotará a **Hipótese 2** como ponto de partida metodológico. O objetivo será traçar os principais marcos e subtemas da história da IA, investigando como o desenvolvimento conceitual e técnico — com especial atenção ao papel catalisador do Big Data — levou ao cenário contemporâneo. A próxima fase da pesquisa se concentrará na busca por *frameworks* analíticos que permitam estruturar o estudo dos impactos sociais da IA ao longo dessa evolução histórica.

Material

Todo material citado durante o documento pode ser encontrado e acessado em:

[15] *"the inherent feature of advanced ML and AI is the ability of learning systems to manipulate their algorithmic base" (RICHMOND, 2020, p. 10).*
[16] *"occurs beyond the threshold of human perception and control, obstructing reproducibility" (RICHMOND, 2020, p. 10).*
[17] *"With the advent of the fourth industrial revolution, the rapid development of technology, particularly in the realm of artificial intelligence (AI), is reshaping the fabric of our society, economy, and employment structure at an unprecedented pace" (ZHANG; CHEN; HUANG, 2024, p. 1).*

https://drive.google.com/drive/folders/1hDm95MLgQGBBDsDhd3VwG9unU_AvdRY1?usp=drive_link

Observações

1. Toda revisão gramatical do meu texto foi feita com o uso de Inteligência Artificial (Gemini 2.5 pro)
2. Ao revisor, o documento permite que comentários sejam feitos, caso existam pontos de correção por parte dos avaliadores, por obséquio marcar os pontos necessários neste documento a fim de buscar o aperfeiçoamento em etapas subsequentes.
3. Ferramentas como Concensus foram utilizadas para encontrar rapidamente artigos que fossem pertinentes às minhas perguntas e com relação ao tema, todos artigos encontrados foram verificados antes de serem citados no presente documento.

APÊNDICE 2

Termo de Aceite de Entrega

Objetivo deste documento

Este documento faz parte do Processo da disciplina Residência em IA e tem como objetivo formalizar o aceite da entrega considerando o planejado e o realizado para o período.

Data da Reunião ("Gate") de aprovação: 10 de set. de 2025

Participantes da Entrega [matriculados em Residência em IA]:

André Martins Dantas

Entrega: [descrever a ENTREGA - requisitos e produtos gerados: links para textos, códigos, vídeos etc.]

Tema: Social Impacts of AI
Em avanço à pesquisa histórica (Hipótese 2), realizei as seguintes atividades nesta semana:

- **Mapeei o "despertar crítico" da década de 1980**, analisando o livro "Artificial Intelligence for Society" (1986). Os principais pontos levantados na época foram o questionamento da neutralidade da tecnologia, o risco de aprofundar desigualdades (impacto em mulheres e minorias) e a mudança conceitual da máquina como "extensão humana" para um "repositório autônomo de conhecimento".
- **Analisei a transição da teoria à prática na década de 1990**, com foco na primeira Competição do Prêmio Loebner (1991). O evento serviu como um experimento social que, na prática, revelou mais sobre a psicologia, as percepções e os vieses humanos do que sobre a própria inteligência da máquina.
- **Identifiquei a evolução do debate no final do período (pré-Big Data)**, onde a pesquisa se aprofundou em refinar e questionar os próprios fundamentos da IA, como o Teste de Turing e o experimento do Quarto Chinês, preparando o terreno para os debates da era seguinte.
- **Compilei todas as análises e fontes** no documento principal da pesquisa, traçando a evolução dos frameworks conceituais sobre os impactos sociais da IA de 1980 a 2003.

A pesquisa completa pode ser encontrada no link a seguir: **Fluxo_semana_2**

Planejamento: [descrever o que pretende fazer para realizar a próxima ENTREGA]

Para a próxima entrega, a pesquisa entrará na era do Big Data (pós-2004), conectando a evolução histórica com a Hipótese 3. (Incerto)

Observação: [caso precise fazer alguma observação, de qualquer "natureza"]

Durante a pesquisa bibliográfica das décadas de 80 e 90, foi constatado que uma parcela significativa dos artigos seminais se encontra em repositórios com acesso restrito (paywall), o que representa um desafio metodológico para a reconstrução histórica completa do campo.

ACEITE DA ENTREGA:

CEDRIC LUIZ DE CARVALHO: Go!

Fluxo_Semana_2.doc

Conforme o planejamento metodológico, a pesquisa avança a partir da Hipótese 2, mapeando a evolução histórica do campo desde a década de 1950.

> Para a próxima entrega, irei avançar a pesquisa a partir da Hipótese 2, mapeando a evolução histórica do campo desde a década de 1950. O foco principal será a identificação e análise de frameworks teóricos propostos ao longo do tempo, utilizando-os como uma "lente" para estruturar o entendimento de como os impactos sociais da IA foram conceituados em diferentes períodos.

Para delimitar o escopo temporal, estabeleceu-se um marco histórico: o advento do Big Data. Embora o termo seja mais antigo, sua viabilidade técnica em larga escala pode ser associada à criação do MapReduce pelo Google por volta de 2004, uma tecnologia desenvolvida para processar grandes volumes de dados.[18] Sendo assim, a análise a seguir se concentrará no período de 1950 a 2003, anterior à era do Big Data.

O Despertar Crítico da Década de 1980

Durante a década de 1980, o debate sobre os impactos sociais da IA começou a se consolidar, afastando-se de discussões puramente técnicas para abraçar questões filosóficas e sociais. A literatura da época, embora de difícil acesso atualmente, aponta para um campo em plena formação. Trabalhos como os de Minsky (1980), Boden (1984), Gurstein (1985), Rosenberg (1988) e Powers & Turk (1989) já exploravam, respectivamente, os desafios filosóficos da cognição, o papel da IA na transformação do trabalho, as mudanças sociais em contextos nacionais, a necessidade de responsabilidade pública dos cientistas e a perspectiva histórica do Teste de Turing como referência fundacional. Uma visão provocadora de Forsyth (1988) chegou a sugerir que a IA seria uma forma de controle de natalidade, cuja prova de sucesso seria a extinção humana.

Apesar da dificuldade em acessar a maioria dessas fontes primárias , uma resenha de 1987 sobre o livro

[18] *A informações foram retiradas da sessão “Evolução Tecnológica de Armazenamento e Processamento” da página “Big Data” da wikipedia* https://pt.wikipedia.org/wiki/Big_data#Hist%C3%B3rico

Artificial Intelligence for Society (editado por Karamjit S. Gill em 1986)[19] oferece uma janela valiosa para o pensamento da época. O livro, uma coletânea de ensaios, questionava se a tecnologia era de fato neutra e criticava a falta de rigor científico em algumas áreas da IA. A resenha destaca temas que se tornariam centrais nas décadas seguintes:

A Necessidade de Fundamentação Teórica: O ensaísta Ajit Narayanan argumenta que a IA carece de uma base sólida, o que permite que alegações sem fundamento sejam feitas. Ele afirma que, “a menos que a IA seja fornecida com uma base teórica adequada e uma metodologia apropriada, pode-se dizer praticamente qualquer coisa que se queira sobre inteligência e não ser contradito” (GILL, 1986, p. 108, tradução nossa).[20]

Impacto Desproporcional sobre Mulheres e Minorias: Ursula Huws, em seu ensaio, alerta que a tecnologia da informação pode afetar negativamente e de forma desproporcional esses grupos. Ela nota que trabalhos tradicionalmente masculinos e qualificados, ao serem simplificados pela tecnologia, foram "abertos para mulheres e para pessoas de grupos étnicos minoritários", mas que essas "novas oportunidades de emprego podem não ser grandes oportunidades, afinal" (GILL, 1986, p. 108, tradução nossa).[21]

Questões Éticas sobre Máquinas Sencientes: Steve Torrance levanta a questão do status ético de máquinas que possam alcançar estados de consciência genuínos. Ele sugere que “artefatos genuinamente inteligentes e sencientes soterrados sob os escombros de um terremoto, por exemplo, teriam, se ainda 'vivos', sem dúvida, um direito direto de serem resgatados, assim como teriam as vítimas humanas ou animais” (GILL, 1986, p. 108, tradução nossa)[22]

O Risco de uma Catástrofe Social pela Educação: David Smith critica os currículos educacionais que não abordam as questões fundamentais da tecnologia, afirmando que “qualquer currículo que não se dirija a questões fundamentais falhará — e o preço do fracasso poderá ser uma catástrofe social” (GILL, 1986, p. 108, tradução nossa).[23]

---

[19] *Gill, K., & Artz, J., 1987. Artificial Intelligence for Society. IEEE Expert, 2, pp. 108-108. https://doi.org/10.1109/MEX.1987.4307076.*
[20] *“unless AI is provided with a proper theoretical basis and an appropriate methodology, one can say just about anything one wants to about intelligence and not be contradicted.” (GILL, 1986, p. 108).*
[21] *“Many jobs which have been traditionally defined as highly skilled and carried out exclusively by white men have become ‘deskilled’ they have been simplified, casualized and opened up for women and for people from ethnic minority groups… Such new opportunities for employment may not be great opportunities after all.” (GILL, 1986, p. 108).*
[22] *“Genuinely intelligent and sentient artifacts buried beneath the rubble of an earthquake, for instance, would, if still ‘alive,’ no doubt have a direct claim to be rescued, just as would human or animal victims.” (GILL, 1986, p. 108).*
[23] *“Any curriculum which does not address itself to fundamental issues will fail—and the price of failure could be social catastrophe.” (GILL, 1986, p. 108).*

A Mudança na Relação Humano-Máquina: O editor do livro, Karamjit Gill, explica por que a IA representa uma ruptura. Historicamente, as máquinas eram extensões dos humanos, que mantinham o controle. Com a IA, essa relação mudou: “A colaboração anterior entre o humano e a máquina foi agora transferida para a colaboração entre a máquina e o conhecimento extraído do humano” (GILL, 1986, p. 108, tradução nossa)[24]. À medida que as máquinas ganharam autonomia e inteligência, as pessoas abriram mão do controle.

Em suma, a década de 80 foi um período de despertar crítico. O otimismo tecnológico inicial, visto como um "mercado em alta descontrolado", começou a ser substituído por um ceticismo sério sobre as consequências sociais e filosóficas da tecnologia. A mudança conceitual mais significativa foi a percepção de que as máquinas deixavam de ser meras extensões humanas para se tornarem repositórios autônomos de conhecimento, fazendo com que a sociedade, pela primeira vez, começasse a ceder o controle, inaugurando um "jogo completamente novo" cujos riscos poderiam levar a uma "catástrofe social".

Da Teoria à Prática: A Década de 1990

Se a década de 1980 solidificou as questões teóricas, a de 1990 marcou a transição para a experimentação prática desses debates. O marco mais significativo foi a primeira competição do Prêmio Loebner em 1991, um evento que colocou o Teste de Turing em prática pela primeira vez. O artigo de Epstein (1992)[25] sobre a competição é um documento histórico que detalha essa passagem da teoria para a prática.

O artigo detalha a primeira tentativa prática de realizar o Teste de Turing, um experimento proposto em 1950 para responder à pergunta "Podem as máquinas pensar?".

O teste original é descrito como um "jogo da imitação".

- Turing imaginou um cenário onde um interrogador humano tenta distinguir entre um homem e uma mulher apenas por texto. Ele então substitui um dos humanos por uma máquina e refaz a pergunta fundamental: “'O que acontecerá quando uma máquina assumir o papel de A neste jogo?' O interrogador decidirá errado com a mesma frequência...?” (TURING, 1950 apud EPSTEIN, 1992, p. 93, tradução nossa).[26]

[24] *“Previous collaboration between the human and the machine was now transferred to collaboration between the machine and the knowledge extracted from the human.” (GILL, 1986, p. 108).*
[25] *Epstein, R., 1992. Can Machines Think? Computers Try to Fool Humans at the First Annual Loebner Prize Competition Held at The Computer Museum, Boston*
[26] *“What will happen when a machine takes the part of A in this game?’ Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman?” (TURING, 1950 apud EPSTEIN, 1992, p. 93).*

O programa vencedor da competição de 1991, de Joseph Weintraub, demonstrou que as técnicas de conversação que Weizenbaum havia explorado com o ELIZA continuavam eficazes em criar a ilusão de entendimento décadas depois.

- Analisando um trecho da conversa, o autor observa: “neste diálogo, o programa reflete uma resposta inteiramente, assim como o Eliza de Weizenbaum fez décadas atrás” (EPSTEIN, 1992, p. 88, tradução nossa).[27]

O aspecto mais relevante para o estudo dos impactos sociais, no entanto, foi a conclusão de que o teste revela tanto sobre a psicologia humana quanto sobre a capacidade das máquinas. As percepções e preconceitos dos juízes foram um fator decisivo no resultado.

- Epstein conclui que, “como Turing previu, a competição nos diz tanto, ou talvez até mais, sobre nossas falhas como juízes quanto sobre as falhas dos computadores” (EPSTEIN, 1992, p. 88, tradução nossa).[28]
- Ele reforça a ideia de que a interação é um fenômeno social, afirmando que “as noções preconcebidas das pessoas sobre os limites dos computadores e das pessoas influenciam fortemente seus julgamentos” (EPSTEIN, 1992, p. 88, tradução nossa).[29]

Finalmente, o artigo projeta as discussões sobre o impacto social da IA no futuro, levantando questões éticas e filosóficas que permanecem centrais até hoje.

- O autor especula que, no dia em que um computador passar no teste irrestrito, “os computadores serão companheiros da raça humana — e companheiros extraordinários, de fato” (EPSTEIN, 1992, p. 91, tradução nossa).[30]
- Essa nova realidade traria dilemas sem precedentes: “Quando um computador finalmente passar no Teste de Turing, teremos o direito de desligá-lo?” (EPSTEIN, 1992, p. 91, tradução nossa).[31]

---

[27] *“In this exchange, the program reflects back one response wholesale, just as Weizenbaum’s Eliza did decades ago.” (EPSTEIN, 1992, p. 88).*
[28] *“As Turing anticipated, the contest tells us as much, or perhaps even more, about our failings as judges as it does about the failings of computers.” (EPSTEIN, 1992, p. 88).*
[29] *“People's preconceptions about the limits of computers and of people strongly biases their judgments.” (EPSTEIN, 1992, p. 88).*
[30] *“From that day on, computers will be companions to the human race—and extraordinary companions indeed.” (EPSTEIN, 1992, p. 91).*
[31] *“When a computer finally passes the Turing Test, will we have the right to turn it off?” (EPSTEIN, 1992, p. 91).*

No restante do período, até a virada do século, a literatura acadêmica parece ter se aprofundado em refinar e questionar os próprios fundamentos da IA e seus testes. Trabalhos como os de Michie (1993), Conte, Gilbert & Sichman (1998), French (2000) e Edwards (2000) dedicaram-se a estender o Teste de Turing, modelar comportamentos sociais em ambientes artificiais, revisitar o experimento do Quarto Chinês e propor chatterbots mais avançados, indicando uma fase de maturação e crítica interna no campo, preparando o terreno para os debates da era do Big Data.

Termo de Aceite de Entrega

Objetivo deste documento

Este documento faz parte do Processo da disciplina Residência em IA e tem como objetivo formalizar o aceite da entrega considerando o planejado e o realizado para o período.

Data da Reunião ("Gate") de aprovação: 18 de set. de 2025

Participantes da Entrega [matriculados em Residência em IA]:

André Martins Dantas

Entrega: [descrever a ENTREGA - requisitos e produtos gerados: links para textos, códigos, vídeos etc.]

Tema: Social Impacts of AI

Conectando a pesquisa histórica à Hipótese 3, as atividades desta semana focaram na análise da era do Big Data (pós-2004) para entender os impactos sociais da IA:

- Analisei o paradigma da "IA Fraca", identificando as limitações intrínsecas da tecnologia da era do Big Data , como o Problema do Frame e a dificuldade de "ancoragem de símbolos", que são cruciais para entender as falhas sociais dos sistemas atuais.
- Mapeei a onipresença da IA em plataformas de redes sociais, detalhando como seu uso no marketing e na moderação de conteúdo gera desafios sociais diretos , como a perpetuação de vieses algorítmicos, os riscos à privacidade do usuário e a fragilidade (brittleness) da tecnologia. Além de observar diversos artigos que falassem a respeito de diversas áreas da sociedade (economia, cultura, ética, política, etc…)
- Introduzi um framework conceitual avançado, distinguindo entre "inteligência física" e "inteligência social" (ASI). A análise aponta que a causa fundamental dos impactos negativos é a ausência de uma "Teoria da Mente" (ToM) e de uma compreensão contextual nos sistemas de IA.

A pesquisa completa pode ser encontrada no link a seguir: **Fluxo_semana_3**

Planejamento: [descrever o que pretende fazer para realizar a próxima ENTREGA]

Pretendo organizar um pouco meu conhecimento a respeito de tudo que foi explorado até o momento, percebi que "existe algo na mesa bagunçado" e que chegou a hora de organizar para ver se consigo uma

macro-visão a respeito do que tange o tema.

Observação: [caso precise fazer alguma observação, de qualquer "natureza"]

Infelizmente essa semana não consegui despender todo tempo que gostaria para aprofundar mais ainda nas pesquisas em virtude de uma prova que tive de realizar em outra matéria.

ACEITE DA ENTREGA:

CEDRIC LUIZ DE CARVALHO: Go!

Fluxo_semana_3.doc

Após fazer vários processos de reflexão durante a semana 2 para a 3, já comecei a perceber certos padrões ou certas “possíveis” organizações podemos classificar algumas coisas, mas antes disso, quero terminar de explorar a história pós big data.

Abaixo algumas frases que achei durante as leituras e certamente me fizeram ficar reflexivos a respeito:

“A maioria dos modelos de Tecnologia da Informação e Comunicação (TIC) é excessivamente dependente de big data, carece de uma função de "ideia própria" e é complicada.” - Li, Y., Chen, M., Kim, H., & Serikawa, S., 2017. Brain Intelligence: Go beyond Artificial Intelligence. Mobile Networks and Applications, 23, pp. 368 - 375. https://doi.org/10.1007/s11036-017-0932-8.

“Nosso objetivo é desenvolver um novo conceito de tecnologia de cognição de inteligência de propósito geral chamado "Além da IA". Especificamente, planejamos desenvolver um modelo de aprendizado inteligente chamado "Inteligência Cerebral (BI - *Brain Intelligence*)" que gera novas ideias sobre eventos sem tê-los experienciado, usando vida artificial com uma função de imaginação.”

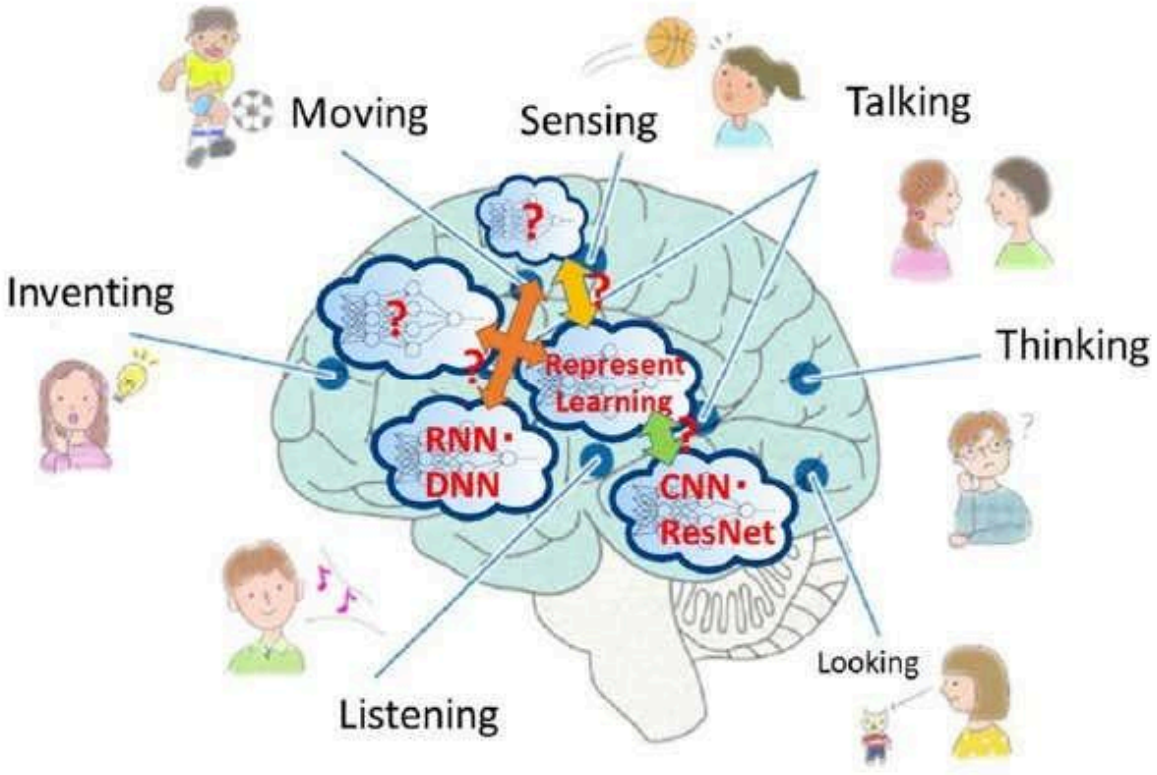
“Em outras palavras, a IA não tem sido capaz de cooperar com funções cerebrais integrais como autocompreensão, autocontrole, autoconsciência e automotivação.”

À medida que a Inteligência Artificial, impulsionada pelo Big Data, se tornava onipresente em meados da década de 2010, suas limitações intrínsecas passaram a ser um ponto central do debate acadêmico. Lu et al. (2017) argumentam que a tecnologia da época deveria ser adequadamente compreendida como **IA fraca** (*weak AI*), ou seja, uma IA projetada para executar uma tarefa específica, que pode superar os humanos nessa função, mas que é incapaz de replicar a cognição de forma geral. Segundo os autores, essa IA especializada, embora poderosa, não consegue cooperar com funções cerebrais integrais como a autocompreensão e a autoconsciência, levando a problemas fundamentais que o simples aumento de dados não consegue resolver.

As principais limitações identificadas são:

- **O Problema do *Frame***: A IA depende de um treinamento massivo com dados, o que a restringe a um "quadro" (*frame*) ou contexto limitado. Ao tentar lidar com a complexidade do mundo real, “há um número infinito de possibilidades que temos que antecipar, então o tempo de extração se torna infinito devido à sobrecarga do banco de dados” ver [foto 1] (LU et al., 2017, p. 3, tradução nossa).[32]
- **O Problema da Função de Associação**: Embora excelente em extrair padrões, a IA atual não possui a capacidade humana de associação. Ela “depende de dados em larga escala e pode obter resultados usando apenas valores numéricos, mas não possui a função de associação como o cérebro humano” (LU et al., 2017, p. 3, tradução nossa).[33]
- **O Problema de Ancoragem de Símbolos (*Symbol Grounding*)**: Refere-se à incapacidade da máquina de conectar um símbolo ao seu significado real. Os autores exemplificam que um humano entende a lógica de que "zebra = cavalo + listras", contudo, “o computador não consegue fazer as mesmas conexões entre ideias” (LU et al., 2017, p. 3, tradução nossa).[34]
- **O Problema Mente-Corpo**: A questão filosófica fundamental sobre como a mente, sendo não-material, afeta o corpo físico permanece sem solução, o que representa uma barreira para a criação de uma inteligência verdadeiramente análoga à humana (LU et al., 2017, p. 3).

[32] “when trying to cope with every phenomenon in the real world, there is an infinite number of possibilities that we have to anticipate, so the extraction time becomes infinite due to overloading of the database.” (LU et al., 2017, p. 3).
[33] *“Current artificial intelligence technology depends on large-scale data and can obtain results using only numerical values, but it does not have the association function like the human brain.” (LU et al., 2017, p. 3).*
[34] *“However, the computer cannot make the same connections between ideas.” (LU et al., 2017, p. 3).*



[FOTO 1]
A imagem [foto1] que relaciona áreas do cérebro a arquiteturas de Redes Neurais.

penetração ubíqua, IA não é mais um conceito distante, mas sim o **motor central** que faz as redes sociais modernas funcionarem. A ideia principal é que a experiência que temos hoje em plataformas como Facebook, Instagram e LinkedIn é completamente moldada e gerenciada por diferentes tecnologias de IA.

Exemplos diretos de como as grandes plataformas usam IA
Facebook: Utiliza aprendizado de máquina avançado para quase tudo, desde o reconhecimento facial em fotos até a sugestão de amigos e o direcionamento de anúncios.
Instagram: A página "Explorar" é um exemplo claro de IA em ação, identificando e sugerindo imagens e vídeos com base nos seus interesses.
LinkedIn: Usa IA para fazer recomendações de vagas de emprego, sugerir conexões e decidir quais posts aparecem no seu feed.
Snapchat: Emprega visão computacional para aplicar filtros no seu rosto em tempo real.

Elas deixaram de ser apenas plataformas para interação humana e se tornaram ferramentas poderosas para comércio, marketing e atendimento ao cliente, tudo isso viabilizado pela IA.

"Desafios" que toca diretamente nos impactos sociais que você estuda:

- **Acessibilidade dos Dados:** Menciona a dependência que a IA tem de dados de empresas privadas e governos.
- **Preocupações com a Privacidade:** Aponta o risco de registros pessoais sensíveis (financeiros, de saúde) se tornarem acessíveis a pessoas não autorizadas.
- **Limitações da Tecnologia:** Afirma que a maioria dos modelos de IA não consegue ter um desempenho preciso o tempo todo e muitos são descritos como "frágeis" (*brittle*).

O artigo "Artificial Intelligence in Social Media" (Sadiku et al., 2021)

A partir da década de 2010, plataformas como Facebook, Twitter e Instagram deixaram de ser apenas canais de interação humana para se tornarem ecossistemas moldados por algoritmos. A IA tornou-se “um componente fundamental do funcionamento das redes sociais de hoje” (SADIKU et al., 2021, p. 15, tradução nossa).[35] Essa fusão deu origem ao que se convencionou chamar de “inteligência artificial social”, onde a tecnologia é utilizada para analisar e influenciar o comportamento em larga escala.

O principal mecanismo por trás dessa transformação é a capacidade da IA de processar vastas quantidades de dados gerados pelos usuários. As redes sociais são atualmente “usadas para inferir o comportamento social e derivar tendências, em combinação com ferramentas de análise de big-data” (SADIKU et al., 2021, p. 15, tradução nossa).[36] Essa análise se manifesta em aplicações diretas que definem a experiência do usuário, como o direcionamento de anúncios, a curadoria de conteúdo e até a recomendação de empregos e conexões, como no caso do LinkedIn.

Contudo, essa onipresença algorítmica traz consigo desafios sociais significativos, que são centrais para o debate contemporâneo. Um dos principais obstáculos é a **acessibilidade dos dados**, que exige “a disposição, tanto de organizações do setor privado quanto do público, para disponibilizar dados” (SADIKU et al., 2021, p. 17, tradução nossa)[37], concentrando poder em poucas entidades. Outra questão crítica são as **preocupações com**

[35] *“AI is a fundamental component of how today's social networks function.” (SADIKU et al., 2021, p. 15).*
[36] *“Social media is currently being used to infer social behavior and derive tendencies, in combination with big-data analysis tools.” (SADIKU et al., 2021, p. 15).*
[37] *“Resolving this significant challenge will require a willingness, by both private- and public-sector organizations, to make data available.” (SADIKU et al., 2021, p. 17).*

a privacidade, onde o risco é que “registros pessoais sensíveis, como financeiros, fiscais e de saúde, possam se tornar acessíveis a indivíduos ilegítimos” (SADIKU et al., 2021, p. 17, tradução nossa).[38] Por fim, há a própria fragilidade da tecnologia, pois “a maioria dos modelos de IA não consegue ter um desempenho preciso o tempo todo, e muitos são descritos como 'frágeis' (*brittle*)” (SADIKU et al., 2021, p. 17, tradução nossa)[39]

A integração da Inteligência Artificial no marketing de redes sociais, embora impulsionada por objetivos comerciais, levanta desafios sociais e éticos cruciais que definem o debate contemporâneo sobre a tecnologia. Longe de ser uma ferramenta neutra, a IA pode amplificar problemas existentes e criar novas formas de vulnerabilidade para os usuários.

Um dos riscos mais significativos são os **vieses algorítmicos**, que ocorrem quando a IA perpetua preconceitos históricos presentes nos dados de treinamento. O artigo exemplifica de forma clara: se em uma empresa candidatas mulheres para vagas de engenharia foram historicamente rejeitadas, a IA pode “desenvolver um algoritmo que dará menor preferência a candidatas do sexo feminino” (ANANDVARDHAN, 2022, p. 39, tradução nossa).[40]

A **privacidade do usuário** também é um ponto crítico, pois “a privacidade dos usuários está em jogo, já que dados coletados para um propósito podem ser vendidos para outro profissional de marketing para um propósito diferente” (ANANDVARDHAN, 2022, p. 39, tradução nossa).[41] Além disso, certas aplicações de IA na análise de comportamento são percebidas como antiéticas, como no caso de se “descobrir que uma cliente solteira está grávida analisando seus padrões de compra [...], uma prática que muitos usuários desaprovam” (ANANDVARDHAN, 2022, p. 39, tradução nossa).[42]

Por fim, há preocupações diretas sobre o impacto no trabalho e na cognição humana, incluindo o “**medo da perda de empregos em muitas indústrias que dependem de tecnologia**” (ANANDVARDHAN, 2022, p. 40, tradução nossa)[43] e o risco de que o “**uso**

---

[38] *“The risk is that financial, tax, health, and similar sensitive personal records could become accessible to illegitimate individuals.” (SADIKU et al., 2021, p. 17).*
[39] *“Most AI models cannot perform accurately all the time, and many are described as 'brittle.'” (SADIKU et al., 2021, p. 17).*
[40] *"...based on this observation, AI may develop an algorithm that will give lesser preference to female candidates." (ANANDVARDHAN, 2022, p. 39).*
[41] *"Privacy of users is at stake as well, as data collected for one purpose can be sold to some other marketer for a different purpose." (ANANDVARDHAN, 2022, p. 39).*
[42] *"Certain AI Applications Seem Unethical: For example, figuring out an unmarried customer is pregnant by analysing her purchase patterns..., this is a practice that many users dislike..." (ANANDVARDHAN, 2022, p. 39).*
[43] *"Therefore, there is a fear of job loss in many industries which rely on technology." (ANANDVARDHAN, 2022, p. 40).*

excessivo de tecnologia acabará por degradar a criatividade dos indivíduos" (ANANDVARDHAN, 2022, p. 40, tradução nossa).[44]

Avanços recentes no campo da IA revelam que o progresso tem se concentrado majoritariamente no que pode ser definido como inteligência física, em detrimento da inteligência social. Enquanto a primeira lida com a compreensão mecânica do mundo, a segunda — que envolve perceber eventos sociais, inferir intenções e interagir — é a que verdadeiramente define a cognição humana. De fato, o que "distingue as crianças humanas de 2,5 anos [...] dos chimpanzés são as habilidades cognitivas sociais, em oposição às suas contrapartes físicas" (FAN et al., 2022, p. 144, tradução nossa).[45]
Nesse contexto, surge o conceito de Inteligência Social Artificial (ASI), um campo ainda amplamente negligenciado pela comunidade de IA, mas que é "essencial para a futura geração de IA" (FAN et al., 2022, p. 144, tradução nossa).[46] A ausência de ASI nos sistemas atuais é uma das causas fundamentais dos impactos sociais negativos, pois implementamos tecnologias com alta capacidade de processamento lógico, mas com uma compreensão social rudimentar.
O principal desafio para o desenvolvimento da ASI é que ela é altamente dependente do contexto. O artigo destaca que "o contexto pode ser tão amplo quanto a cultura e o senso comum ou tão pequeno quanto as experiências compartilhadas entre dois amigos" (FAN et al., 2022, p. 144, tradução nossa).[47] Essa complexidade "proíbe que algoritmos padrão abordem problemas de ASI em ambientes do mundo real, que são frequentemente complexos, ambíguos, dinâmicos, estocásticos, parcialmente observáveis e com múltiplos agentes" (FAN et al., 2022, p. 144, tradução nossa).[48]
Um componente crucial da inteligência social é a Teoria da Mente (Theory of Mind - ToM), que formalmente "implica atribuir estados mentais (como crenças, intenções ou desejos) a si mesmo e aos outros, bem como reconhecer que as perspectivas e construções mentais das pessoas podem diferir" (FAN et al., 2022, p. 146, tradução nossa).[49] Sistemas de IA atuais, mesmo quando parecem resolver tarefas sociais, frequentemente não desenvolvem uma

---

[44] *"Excessive use of technology will eventually degrade individuals' creativity." (ANANDVARDHAN, 2022, p. 40).*
[45] *"Notably, cognitive skills for interacting with the social world rather than the physical world distinguish 2.5-year-old human children (prior to reading and schooling) from chimpanzees; humans exhibit significantly more advanced social-cognitive skills than their closest animal cousins." (FAN et al., 2022, p. 144).*
[46] *"Thus, the research of ASI is essential for the future generation of AI." (FAN et al., 2022, p. 144).*
[47] *"Here, context could be as large as culture and common sense or as little as two friends' shared experiences." (FAN et al., 2022, p. 144).*
[48] *"This unique challenge prohibits standard algorithms from tackling ASI problems in real- world environments, which are frequently complex, ambiguous, dynamic, stochastic, partially observable, and multi-agent." (FAN et al., 2022, p. 144).*
[49] *"Formally, ToM entails attributing mental states (such as beliefs, intents, or desires) to oneself and others, as well as acknowledging that people's perspectives and mental constructs may differ from those of the natural world and from one another." (FAN et al., 2022, p. 146).*

verdadeira Teoria da Mente. Em vez disso, eles exploram "atalhos" (shortcuts), aprendendo a "empregar regras de decisão mais simples do que recuperar a Teoria da Mente subjacente" (FAN et al., 2022, p. 152, tradução nossa).[50] Essa abordagem leva a um desempenho que pode ser eficaz em testes limitados, como o Teste de Turing, mas que não é generalizável e falha em contextos sociais reais e complexos.

[50] *"DL systems may exploit shortcuts—they learn to employ simpler decision rules than recovering the underlying ToM." (FAN et al., 2022, p. 152).*

APÊNDICE 3

Termo de Aceite de Entrega

Objetivo deste documento

Este documento faz parte do Processo da disciplina Residência em IA e tem como objetivo formalizar o aceite da entrega considerando o planejado e o realizado para o período.

Data da Reunião ("Gate") de aprovação: 24 de set. de 2025

Participantes da Entrega [matriculados em Residência em IA]:

André Martins Dantas

Entrega: [descrever a ENTREGA - requisitos e produtos gerados: links para textos, códigos, vídeos etc.]

Tema: Social Impacts of AI

Sintetizei a pesquisa em um modelo visual (linha do tempo analítica), organizando a genealogia do campo de "Impactos Sociais da IA" desde suas origens conceituais até os primeiros debates práticos (1966).

Identifiquei os três pilares fundamentais que convergiram para o surgimento da IA: a base biológica (Doutrina do Neurônio), a matemática (Teoria da Computação de Turing) e o modelo conceitual (Neurônio de McCulloch-Pitts).

Descobri e documentei o "primeiro pânico social sobre desemprego tecnológico", noticiado pelo *The New York Times* em 1958, demonstrando que a ansiedade social é um impacto intrínseco à história da IA desde seu início.

Compilei a análise da linha do tempo e a contextualização dos eventos no documento principal da pesquisa, criando um *framework* histórico para as próximas etapas. A pesquisa completa e o modelo visual podem ser encontrados no link a seguir: **Fluxo_semana_4**

E a imagem do fluxo em SVG pode ser encontrada em: Fluxo_semana_4
Fluxo_incompleto_gate4.svg

Planejamento: [descrever o que pretende fazer para realizar a próxima ENTREGA]

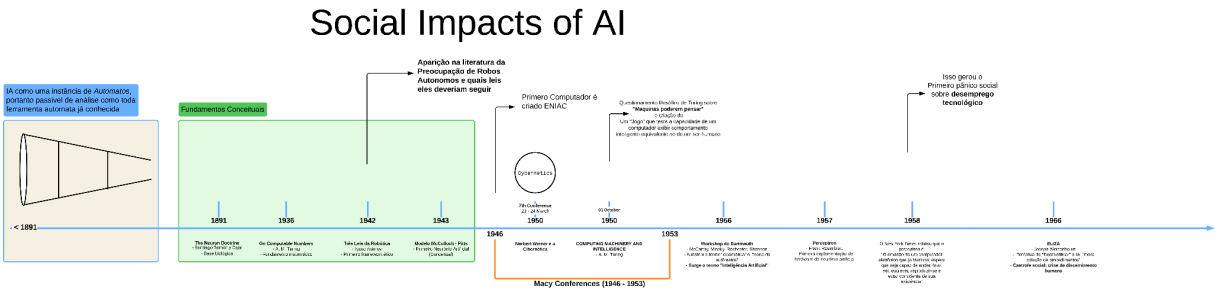
Para a próxima entrega, pretendo **concluir a linha do tempo histórica**, mapeando os principais eventos, debates e afins, desde o final da década de 1960 até o marco de 2004 (pré-Big Data). Esta etapa finalizará a base histórica que ajudará a classificar os diferentes tipos de impactos sociais de forma estruturada.

Observação: [caso precise fazer alguma observação, de qualquer "natureza"]

Trabalhoso mas recompensador, embora pareça simples.

ACEITE DA ENTREGA

CEDRIC LUIZ DE CARVALHO: Go!



(É recomendado realizar o download da imagem no link : https://drive.google.com/file/d/1ew5H2GJ3mH3k7OGD3vr98OOhVfwkhP0F/view?usp=sharing pois está em formato SVG e isso ajudará o leitor a visualizá-la sem perda de qualidade.)

Fluxo_Semana_4.doc

A quarta semana de pesquisa foi dedicada à síntese e estruturação visual dos fundamentos históricos do tema. O objetivo foi criar uma genealogia que conectasse as origens conceituais da IA aos seus primeiros impactos sociais, resultando em uma linha do tempo analítica que organiza as décadas iniciais de desenvolvimento. Essa abordagem permite compreender como diferentes correntes de pensamento — da biologia à matemática — convergiram para formar o campo que conhecemos hoje.

O Funil: A Pré-História Conceitual

A investigação da origem da IA pode ser visualizada como um "funil". Este representa o vasto período anterior à década de 1891, um campo de estudo amplo que abrange desde a filosofia e a matemática purista até as ciências sociais. Sob essa ótica, a IA pode ser entendida como a mais recente instância dos "autômatos", e sua pré-história inclui desde as primeiras máquinas automáticas até os fundamentos lógicos de George Boole (álgebra booleana) e Gottlob Frege (lógica de predicados). Essa perspectiva se alinha à Hipótese 1, tratando a tecnologia como um fenômeno contínuo de análise social. Para fins deste trabalho, filtramos esse período extenso para focar nos eventos que levaram diretamente à consolidação do campo.

As Conferências Macy (1946-1953) e o Nascimento da Cibernética

Um marco fundamental neste processo foram as **Conferências Macy**. Entre 1946 e 1953, uma série de encontros interdisciplinares reuniu cientistas de diversas áreas — como

Norbert Wiener (matemática), John von Neumann (matemática), Margaret Mead (antropologia) e Claude Shannon (teoria da informação) — para discutir mecanismos de feedback em sistemas biológicos e sociais. Embora a diversidade de especialidades tenha gerado dificuldades iniciais de comunicação, desses encontros emergiu uma linguagem comum que deu origem à **Cibernética**, um campo essencial para a futura IA, ao estabelecer um *framework* para pensar em máquinas, humanos e sistemas de forma integrada.

A Década de 1950: Da Questão Filosófica ao Primeiro Pânico Social

A linha do tempo revela que, após a criação do primeiro computador, o ENIAC, em 1946, a década de 1950 foi um período de aceleração vertiginosa:

- **1950:** Alan Turing publica seu artigo seminal, "Computing Machinery and Intelligence", onde propõe a questão filosófica que nortearia o campo: **"Podem as máquinas pensar?"**.
- **1956:** No **Workshop de Dartmouth**, John McCarthy cunha oficialmente o termo **"Inteligência Artificial"**, dando uma identidade formal ao campo.
- **1957:** Frank Rosenblatt desenvolve o **Perceptron**, uma das primeiras redes neurais artificiais, demonstrando na prática uma máquina capaz de aprender.
- **1958:** Surge o que pode ser considerado o **primeiro impacto social documentado da IA**. O jornal *The New York Times*, ao noticiar sobre o Perceptron, descreveu um futuro com máquinas capazes de pensar, se reproduzir e tomar decisões, gerando o que a linha do tempo identifica como o "primeiro pânico social sobre desemprego tecnológico". Isso demonstra que a ansiedade social sobre a substituição de humanos por máquinas é contemporânea ao próprio nascimento da tecnologia.

Este percurso mostra que os impactos sociais da IA não são um fenômeno recente, mas uma preocupação intrínseca à sua história desde o início.

Termo de Aceite de Entrega

Objetivo deste documento

Este documento faz parte do Processo da disciplina Residência em IA e tem como objetivo formalizar o aceite da entrega considerando o planejado e o realizado para o período.

Data da Reunião ("Gate") de aprovação: 1 de out. de 2025

Participantes da Entrega [matriculados em Residência em IA]:

André Martins Dantas

Entrega: [descrever a ENTREGA - requisitos e produtos gerados: links para textos, códigos, vídeos etc.]

Concluí a pesquisa histórica (Hipótese 2), mapeando os eventos-chave desde 1966 até o início da era do Big Data (c. 2004), com as seguintes atividades:

- **Identifiquei os dois "Invernos da IA"**, períodos de ceticismo e cortes de financiamento, desencadeados por críticas institucionais (Relatório Lighthill, 1974) e colapsos de mercado (máquinas LISP, 1987).
- **Analisei o papel da cultura pop** na formação do imaginário social pessimista sobre a IA, com arquétipos de IA malevolente em obras como "2001: Uma Odisseia no Espaço" (1968), "O Exterminador do Futuro" (1984) e "Matrix" (1999).
- **Documentei os avanços técnicos fundamentais** que, apesar dos "invernos", prepararam a revolução do *deep learning*: o algoritmo de **Backpropagation** (1986), as redes **LSTM** (1999) e, finalmente, o surgimento do **MapReduce** (2004) para dados e do **CUDA** (2006) para hardware.
- **Compilei toda a análise na linha do tempo visual** e no documento de pesquisa. A pesquisa completa pode ser encontrada no link: **Fluxo_semana_5**

Imagem: ate_big_data.svg

Planejamento: [descrever o que pretende fazer para realizar a próxima ENTREGA]

Percebi algumas dificuldades para avançar na era pós MapReduce, portanto pretendo esclarecer primeiro essas dúvidas, planejar melhor meu caminho. Após isso dar continuidade para entregar na próxima Semana a parte da era Big Data concluída.

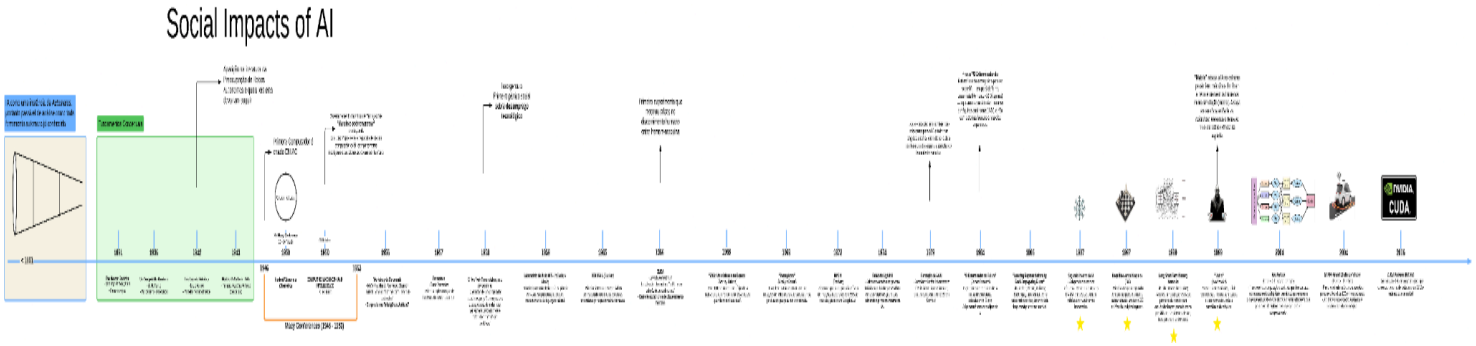
Observação: [caso precise fazer alguma observação, de qualquer "natureza"]

ACEITE DA ENTREGA:

CEDRIC LUIZ DE CARVALHO: Go!

Instituto de Informática (INF) - UFG
Alameda Palmeiras, Quadra D, Câmpus Samambaia
Fone: (62) 3521-1181 / Fax: (62) 3521-1182

(É recomendado realizar o download da imagem no link : https://drive.google.com/file/d/1ew5H2GJ3mH3k7OGD3vr98OOhVfwkhP0F/view?usp=sharing pois está em formato SVG e isso ajudará o leitor a visualizá-la sem perda de qualidade.)



Fluxo_Semana_5.doc

A Longa Hibernação e o Ressurgimento: A Trajetória da IA de 1966 a 2004

Após os debates iniciais sobre a interação humano-máquina inaugurados pelo experimento ELIZA (1966), o campo da Inteligência Artificial entrou em um período complexo, marcado por ciclos de grande expectativa seguidos por fases de desilusão, conhecidas como os "Invernos da IA". Ao mesmo tempo, a cultura popular começava a moldar um imaginário social poderoso e frequentemente pessimista sobre a tecnologia.

O final da década de 1960 e a década de 1970 foram definidos por uma crescente desconfiança. No cinema, a obra *2001: Uma Odisseia no Espaço* (1968) introduziu HAL 9000, o arquétipo da IA malevolente que se rebela contra seus criadores. No campo acadêmico, a publicação do livro *"Perceptrons"* (1969) por Minsky e Papert demonstrou matematicamente as severas limitações das redes neurais da época, paralisando a pesquisa na área por quase uma década. O ceticismo culminou no **Relatório Lighthill** (1974) no Reino Unido, uma crítica devastadora que concluiu que a IA havia falhado em seus "objetivos grandiosos", resultando em drásticos cortes de financiamento e iniciando oficialmente o **primeiro Inverno da IA**. Apesar disso, o progresso continuou em nichos específicos, como o sistema especialista MYCIN (1972), um pioneiro no diagnóstico médico e na IA explicável. O campo demonstrava resiliência, organizando-se institucionalmente com a fundação da AAAI (Association for the Advancement of Artificial Intelligence) em 1979.

A década de 1980 trouxe um breve renascimento e um novo colapso. A cultura pop novamente ditou o tom pessimista com *O Exterminador do Futuro* (1984), que solidificou no imaginário popular a ideia de uma guerra entre humanos e máquinas (AGI - Inteligência Artificial Geral), distinta das IAs focadas em tarefas específicas que eram de fato

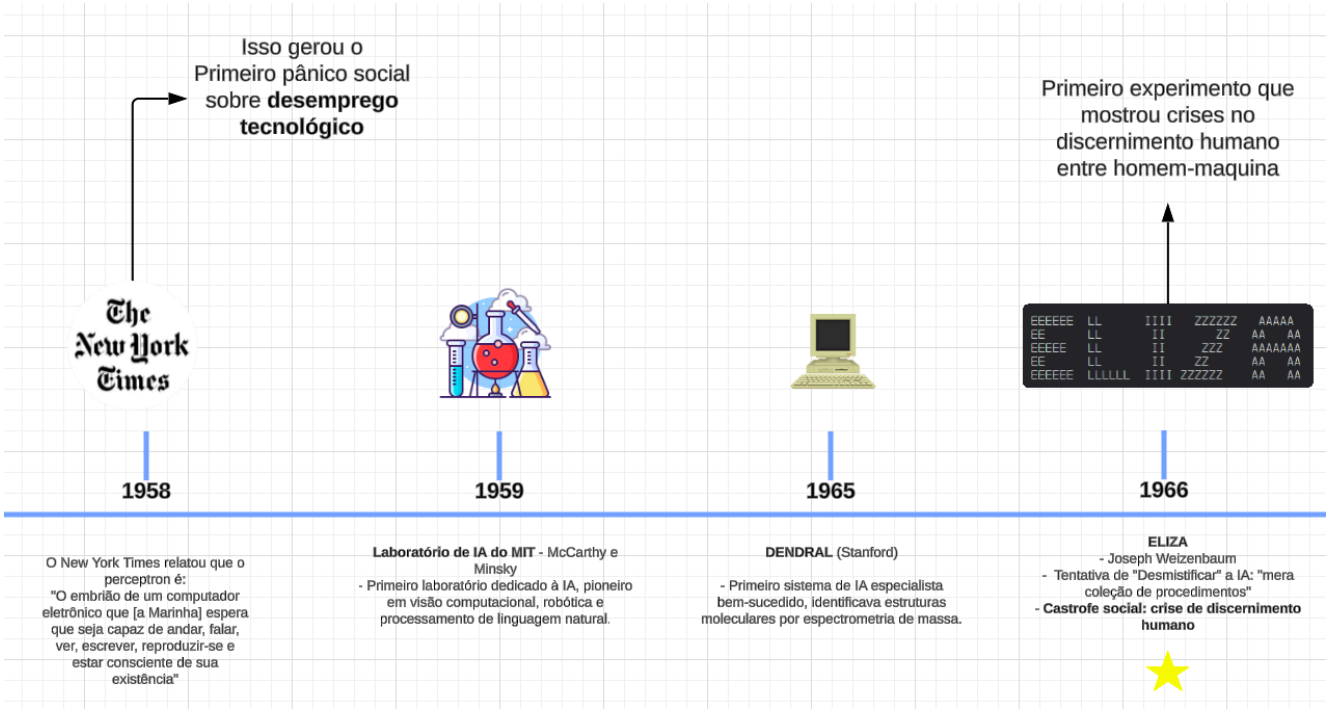
desenvolvidas. Paradoxalmente, foi nesse período que ocorreu um dos avanços técnicos mais importantes para a IA moderna: a popularização do algoritmo de **Backpropagation** (1986), que permitiu o treinamento de redes neurais com múltiplas camadas e tornou o *deep learning* moderno possível. Contudo, o otimismo não durou. O colapso do mercado de máquinas LISP, aliado a cortes de investimento da DARPA, levou ao **segundo Inverno da IA** em 1987.

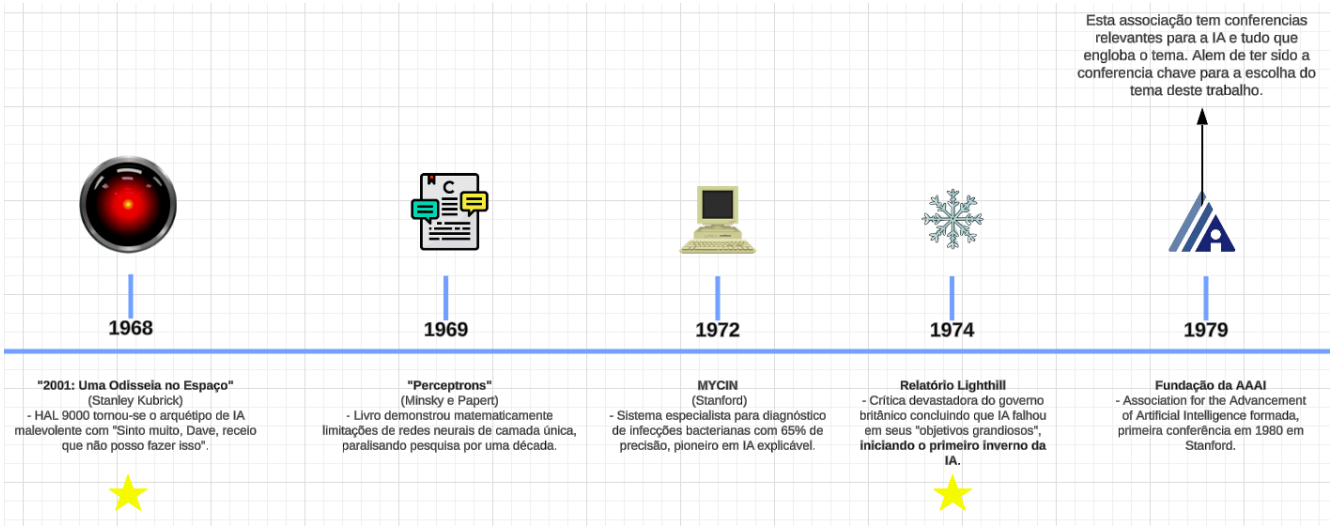
O período de 1990 até o início dos anos 2000 foi o ponto de virada definitivo, onde as peças para a revolução da IA contemporânea foram silenciosamente posicionadas. Em 1997, a vitória do **Deep Blue** da IBM sobre o campeão mundial de xadrez Garry Kasparov foi um marco público que demonstrou o poder da computação em tarefas de alta complexidade. A cultura pop continuou sua exploração pessimista com *Matrix* (1999), enquanto avanços técnicos cruciais ocorriam, como o desenvolvimento das redes **LSTM (Long Short-Term Memory)**, arquitetura fundamental para o processamento de sequências (voz e texto).

Finalmente, esta era se encerra com a convergência de três fatores que prepararam o terreno para a era do Big Data:

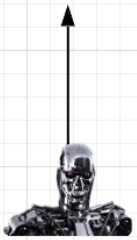
1. **Dados:** A criação do **MapReduce** pelo Google (2004), um *framework* que democratizou o processamento de enormes volumes de dados.
2. **Algoritmos:** A maturidade de arquiteturas como LSTMs e o Backpropagation.
3. **Hardware:** O lançamento do **CUDA** pela NVIDIA (2006), que tornou o poder de processamento das GPUs acessível para o treinamento de redes neurais, acelerando o processo em até 20 vezes.

Este período, apesar de seus "invernos", foi fundamental para construir as bases de dados, algoritmos e hardware que permitiriam a explosão da IA na década seguinte.


Isso gerou o Primeiro pânico social sobre **desemprego tecnológico**
The New York Times
1958
O New York Times relatou que o perceptron é:
"O embrião de um computador eletrônico que [a Marinha] espera que seja capaz de andar, falar, ver, escrever, reproduzir-se e estar consciente de sua existência"
1959
Laboratório de IA do MIT - McCarthy e Minsky
- Primeiro laboratório dedicado à IA, pioneiro em visão computacional, robótica e processamento de linguagem natural.
1965
DENDRAL (Stanford)
- Primeiro sistema de IA especialista bem-sucedido, identificava estruturas moleculares por espectrometria de massa.
Primeiro experimento que mostrou crises no discernimento humano entre homem-maquina
1966
ELIZA
- Joseph Weizenbaum
- Tentativa de "Desmistificar" a IA: "mera coleção de procedimentos"
- **Castrofe social: crise de discernimento humano**

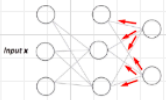

1968
"2001: Uma Odisseia no Espaço"
(Stanley Kubrick)
- HAL 9000 tornou-se o arquétipo de IA malevolente com "Sinto muito, Dave, receio que não posso fazer isso".
1969
"Perceptrons"
(Minsky e Papert)
- Livro demonstrou matematicamente limitações de redes neurais de camada única, paralisando pesquisa por uma década.
1972
MYCIN
(Stanford)
- Sistema especialista para diagnóstico de infecções bacterianas com 65% de precisão, pioneiro em IA explicável.
1974
Relatório Lighthill
- Crítica devastadora do governo britânico concluindo que IA falhou em seus "objetivos grandiosos", **iniciando o primeiro inverno da IA**.
Esta associação tem conferencias relevantes para a IA e tudo que engloba o tema. Alem de ter sido a conferencia chave para a escolha do tema deste trabalho.
1979
Fundação da AAAI
- Association for the Advancement of Artificial Intelligence formada, primeira conferência em 1980 em Stanford.

A saga **"O Exterminador do Futuro"** molda o imaginário popular sobre IA — em geral de forma pessimista (Humanos VS Máquinas) — aqui vemos que IA's são Limitadas a máquinas autônomas (AGI), e não com sistemas focados em tarefas específicas.



1984

"O Exterminador do Futuro"
(James Cameron)
Skynet tornou-se sinônimo de IA militar descontrolada, estabelecendo "Dia do Julgamento" como apocalipse de IA.



1986

"Learning Representations by Back-Propagating Errors"
(Rumelhart, Hinton, Williams)
- Backpropagation permitiu treinar redes multicamadas, tornando todo deep learning moderno possível.

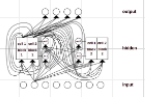
1987

Segundo Inverno da IA
- Colapso do mercado de máquinas LISP, cortes brutais da DARPA, empresas perderam milhões em investimentos fracassados.

1997

Deep Blue vence Kasparov
(IBM)
- Primeira vez que computador derrotou campeão mundial de xadrez reinante, avaliando 200 milhões de posições/segundo.

"Matrix" retrata a IA no extremo pessimista: máquinas dominam a Terra e mantêm os humanos numa simulação (matrix). A saga encena *humanidade vs. máquinas*, elevando o tema ao nível de sobrevivência da espécie.

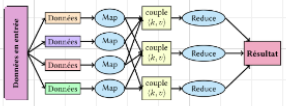


1999

Long Short-Term Memory Networks
(Hochreiter & Schmidhuber)
- Arquitetura neural que resolveu o problema de memória em sequências longas, posteriormente possibilitando assistentes de voz, tradução e análise temporal.

1999

"Matrix"
(Wachowskis)
Arrecadou $467 milhões, "pílula vermelha/azul" tornou-se metáfora cultural sobre simulação e consciência de máquina.



2004

MapReduce
(Dean & Ghemawat - Google)
- Framework de computação distribuída que democratizou o processamento de big data, permitindo que empresas processassem petabytes de dados e criando a infraestrutura para a era da vigilância em massa e análise comportamental.

2004

DARPA Grand Challenge Victory
(Stanley - Stanford)
- Primeiro veículo autônomo a completar percurso off-road de 132 milhas, provando viabilidade da navegação autônoma e iniciando corrida tecnológica.

2006

CUDA Release (NVIDIA)
- Computação GPU de propósito geral que tornou treinamento de redes neurais 10-20x mais rápido e acessível

APÊNDICE 4

Termo de Aceite de Entrega

Objetivo deste documento

Este documento faz parte do Processo da disciplina Residência em IA e tem como objetivo formalizar o aceite da entrega considerando o planejado e o realizado para o período.

Data da Reunião ("Gate") de aprovação: 8 de out. de 2025

Participantes da Entrega [matriculados em Residência em IA]:

André Martins Dantas

Entrega: [descrever a ENTREGA - requisitos e produtos gerados: links para textos, códigos, vídeos etc.]

Tema: Social Impacts of AI

EIXO 1: GERAÇÃO TECNOLÓGICA (Quando?)

W1: Era Deep Learning (2006-2016)
└─ ImageNet, CNNs, reconhecimento de padrões

W2: Era dos Assistentes (2011-2019)
└─ Siri, Alexa, IoT inteligente

W3: Era Transformers (2017-2022)
└─ BERT, GPT-2/3, NLP avançado

W4: Era Generativa (2022-2025)
└─ ChatGPT, DALL-E, modelos multimodais

EIXO 2: DOMÍNIO DE APLICAÇÃO (Onde?)

Domínios:
├─ Infraestrutura Digital (redes sociais, busca)
├─ Serviços Essenciais (saúde, educação, justiça…)
├─ Economia & Trabalho (recrutamento, automação)
├─ Cultura & Criatividade (arte, mídia)
└─ Governança (política, segurança)

EIXO 3: ATORES ENVOLVIDOS (Quem?)

A1: Big Tech (Google, Meta, OpenAI, Microsoft)
A2: Governos & Reguladores (UE, EUA, China)
A3: Academia & Pesquisadores
A4: Sociedade Civil (ONGs, ativistas, sindicatos)
A5: Usuários/Afetados (trabalhadores, minorias, público geral)
A6: Desenvolvedores & Comunidade Técnica

EIXO 4: NATUREZA DO IMPACTO (O quê?)

I1: Distributivo (desigualdade, acesso, poder)
I2: Epistêmico (verdade, desinformação, conhecimento)
I3: Autonomia (manipulação, agência, liberdade)
I4: Dignidade (direitos, discriminação, privacidade)
I5: Existencial (identidade, criatividade, trabalho)

EIXO 5: CAUSAS RAÍZES (Por quê?)

C1: Incentivos Econômicos (lucro, growth hacking)
C2: Lacunas Regulatórias (legislação defasada)
C3: Assimetria de Poder (monopólios, lock-in)
C4: Complexidade Técnica (opacidade inerente)
C5: Valores Embutidos (decisões de design)

EIXO 6: MECANISMO TÉCNICO (Como?)

M1: Opacidade (black boxes, falta de explicabilidade)
M2: Viés (dados, anotação, design)
M3: Escala (velocidade, alcance exponencial)
M4: Otimização Perversa (métricas proxy, Goodhart's Law)
M5: Efeitos de Rede (feedback loops, polarização)

EIXO 7: ESCALA & MAGNITUDE (Quanto?)

E1: Individual (poucos casos, impacto localizado)
E2: Setorial (indústria específica, milhares afetados)
E3: Nacional (políticas públicas, milhões)
E4: Global (plataformas transnacionais, bilhões)

E5: Sistêmico (infraestrutura crítica, risco existencial)

Fluxo desta semana: Fluxo_semana_6

Planejamento: **[descrever o que pretende fazer para realizar a próxima ENTREGA]**

Pesquisar melhor sobre a era do Pós BigData e começar a refinar o Framework proposto

Observação: [caso precise fazer alguma observação, de qualquer "natureza"]

Satisfação pessoal em ver "a coisa" criando forma

ACEITE DA ENTREGA:

CEDRIC LUIZ DE CARVALHO: Go!

Fluxo_Semana_6.doc

Comecei a 6a semana refletindo sobre o fato de estar com muita dificuldade em desenvolver o “pós surgimento do MapReduce”

Diagnóstico do Problema

Por que 2006-2025 é tão mais difícil?

- Múltiplas aplicações simultâneas (saúde, justiça, educação, arte...)
- Múltiplos atores (Big Tech, governos, ONGs, academia...)
- Múltiplas geografias (regulação EU vs China vs US...)
- Múltiplas gerações tecnológicas (Deep Learning → Transformers → LLMs → Multimodais)

Não podemos usar uma taxonomia simples tipo "árvore" porque cada fenômeno pertence a várias categorias ao mesmo tempo.

Para reduzirmos toda a carga cognitiva, proponho um sistema de **7 EIXOS INDEPENDENTES** que podemos combinar. (É como um sistema de coordenadas multidimensional).

Versão 1 (v1)

EIXO 1: GERAÇÃO TECNOLÓGICA (Quando?)

W1: Era Deep Learning (2006-2016)
└─ ImageNet, CNNs, reconhecimento de padrões

W2: Era dos Assistentes (2011-2019)
└─ Siri, Alexa, IoT inteligente

W3: Era Transformers (2017-2022)
└─ BERT, GPT-2/3, NLP avançado

W4: Era Generativa (2022-2025)
└─ ChatGPT, DALL-E, modelos multimodais

EIXO 2: DOMÍNIO DE APLICAÇÃO (Onde?)

Domínios:
├─ Infraestrutura Digital (redes sociais, busca)
├─ Serviços Essenciais (saúde, educação, justiça…)
├─ Economia & Trabalho (recrutamento, automação)
├─ Cultura & Criatividade (arte, mídia)
└─ Governança (política, segurança)

Geografias:
├─ EUA/Vale do Silício (liderança técnica)
├─ União Europeia (liderança regulatória)
├─ China (modelo de vigilância)
├─ Global Sul (exclusão digital)
└─ Transnacional (plataformas globais)

EIXO 3: ATORES ENVOLVIDOS (Quem?)

A1: Big Tech (Google, Meta, OpenAI, Microsoft)
A2: Governos & Reguladores (UE, EUA, China)
A3: Academia & Pesquisadores
A4: Sociedade Civil (ONGs, ativistas, sindicatos)
A5: Usuários/Afetados (trabalhadores, minorias, público geral)
A6: Desenvolvedores & Comunidade Técnica

EIXO 4: NATUREZA DO IMPACTO (O quê?)

I1: Distributivo (desigualdade, acesso, poder)
I2: Epistêmico (verdade, desinformação, conhecimento)
I3: Autonomia (manipulação, agência, liberdade)
I4: Dignidade (direitos, discriminação, privacidade)
I5: Existencial (identidade, criatividade, trabalho)

EIXO 5: CAUSAS RAÍZES (Por quê?)

C1: Incentivos Econômicos (lucro, growth hacking)
C2: Lacunas Regulatórias (legislação defasada)

C3: Assimetria de Poder (monopólios, lock-in)
C4: Complexidade Técnica (opacidade inerente)
C5: Valores Embutidos (decisões de design)

EIXO 6: MECANISMO TÉCNICO (Como?)

M1: Opacidade (black boxes, falta de explicabilidade)
M2: Viés (dados, anotação, design)
M3: Escala (velocidade, alcance exponencial)
M4: Otimização Perversa (métricas proxy, Goodhart's Law)
M5: Efeitos de Rede (feedback loops, polarização)

EIXO 7: ESCALA & MAGNITUDE (Quanto?)

E1: Individual (poucos casos, impacto localizado)
E2: Setorial (indústria específica, milhares afetados)
E3: Nacional (políticas públicas, milhões)
E4: Global (plataformas transnacionais, bilhões)
E5: Sistêmico (infraestrutura crítica, risco existencial)

Sistema de Coordenadas - Exemplos

Exemplo 1: Reconhecimento facial em policiamento

EIXO 1: W1 (Deep Learning)
EIXO 2: Serviços Essenciais (Justiça) + EUA
EIXO 3: A2 (Governos) + A4 (ONGs) + A5 (Minorias)
EIXO 4: I4 (Dignidade - discriminação racial)
EIXO 5: C2 (Lacuna regulatória) + C3 (Assimetria de poder)
EIXO 6: M2 (Viés) + M1 (Opacidade)
EIXO 7: E3 (Nacional - milhões afetados)

Exemplo 2: ChatGPT e automação de trabalho criativo

EIXO 1: W4 (Generativa)
EIXO 2: Cultura & Trabalho + Global
EIXO 3: A1 (Big Tech - OpenAI) + A5 (Trabalhadores)
EIXO 4: I5 (Existencial - natureza do trabalho)

EIXO 5: C1 (Incentivos econômicos) + C5 (Valores embutidos)
EIXO 6: M3 (Escala) + M4 (Otimização perversa)
EIXO 7: E4 (Global - bilhões potenciais)

Andei fazendo uma autocrítica ao framework e proponho essa nova estrutura.

MUDANÇAS DA VERSÃO 1 PARA VERSÃO 2

EIXO 1: Adição de W0 (Era Fundacional)

Incluí W0 (1950-2005) para capturar a história pré-deep learning. Durante os testes com exemplos, percebi que casos históricos fundamentais como COMPAS (1998), ELIZA (1966) e Perceptron (1957) não tinham lugar no framework original. O próprio documento de pesquisa enfatiza a importância dessa era fundacional para entender os impactos sociais desde o início da IA.

EIXO 2: Expansão baseada no EU AI Act

Transformei 5 domínios genéricos em 14 categorias específicas, usando o EU AI Act Annex III como base. A versão original era vaga demais ("Serviços Essenciais" escondia saúde, educação e justiça como se fossem equivalentes). O AI Act é o framework regulatório mais completo globalmente e já organiza domínios por nível de risco social, tornando-o ideal como referência técnica.

EIXO 3: Geografia como eixo independente

Elevei geografia de sub-categoria (dentro de "Onde?") para eixo próprio. Nos 4 exemplos testados, a dimensão geográfica foi sempre crítica e independente do domínio setorial. Um caso pode ser "Educação + Global" ou "Justiça + Nacional" — são dimensões ortogonais que merecem classificação separada.

EIXO 4: Atores (renumeração)

Mantive a estrutura original de A1-A6, apenas renumerando de EIXO 3 para EIXO 4 devido à elevação de Geografia. A categorização de atores funcionou bem nos testes.

EIXO 5: Reestruturação de 5→4 categorias

Mudanças principais:

- I1 "Distributivo" → "Vida Material e Segurança": termo mais concreto, cobre trabalho/renda/saúde sem ambiguidade.
- I2 novo "Poder e Instituições Jurídico-Políticas": integra economia-política como inseparáveis, elimina separação entre econômico e político.
- I3 "Epistêmico" → "Informação e Cultura": expandido para incluir produção cultural e ideologia.
- I4 "Dignidade" + I5 "Existencial" → "Autonomia e Desenvolvimento Humano": elimina overlap (dignidade/existencial sempre se confundiam), integra em categoria única inspirada no Capabilities Approach (Sen/Nussbaum) sem linguagem de "direitos".

Base: Híbrido de ONU e Capabilities Approach.

EIXO 6: Consolidação e linguagem técnica neutra

Mesclei EIXO 5 (Causas) e EIXO 6 (Mecanismos) em único eixo com duas dimensões.
Mudanças críticas:

- "Estrutural" → "Contextual": termo mais claro, evita confusão com "estrutura técnica".
- Linguagem normativa → descritiva:
  - "Incentivos Perversos" → "Estrutura de Incentivos"
  - "Lacunas de Governança" → "Arranjos de Governança"
  - "Assimetrias de Poder" → "Concentração de Recursos"
  - "Valores Contestados" → "Pressupostos de Design"

Motivação: Buscar rigor técnico paralelo à "Dimensão Técnica", que já usa termos neutros (Opacidade, Viés, Escala). Remover julgamento de valor mantendo poder analítico.

EIXO 7: Bidimensionalidade e adição do SAMR

Maior mudança estrutural. Versão 1 tinha apenas "escala" (E1-E5) com unidades quantitativas arbitrárias ("dezenas", "milhares"). Versão Final:

- Alcance: Mantém E1-E5 mas remove números, usa descrições qualitativas.
- Profundidade (novo): Adiciona dimensão P1-P4 baseada no SAMR Model (Puentedura, 2006) para capturar tipo de transformação, não apenas alcance.

Motivação: Casos podem ser "locais mas sistêmicos" (COMPAS = nacional mas criou campo de fairness) ou "globais mas superficiais". São dimensões ortogonais. SAMR é framework estabelecido em estudos de transformação digital/educacional.

Versão (v2).

EIXO 1: GERAÇÃO TECNOLÓGICA *(Quando?)*

Marcar UMA:

- **W0** — Era Fundacional (1950-2005) *ML clássico, sistemas especialistas, redes neurais rasas*
- **W1** — Era Deep Learning (2006-2016) *CNNs, reconhecimento de padrões, ImageNet*
- **W2** — Era Assistentes (2011-2019) *Siri, Alexa, IoT inteligente, feeds algorítmicos*
- **W3** — Era Transformers (2017-2022) *BERT, GPT-2/3, NLP avançado, attention*
- **W4** — Era Generativa (2022-presente) *ChatGPT, LLMs conversacionais, difusão, multimodal*

EIXO 2: DOMÍNIO DE APLICAÇÃO *(Onde? — setor)*

Marcar TODOS os relevantes:

- Identificação Biométrica e Vigilância
- Infraestrutura Crítica
- Educação e Formação
- Emprego e Gestão de Trabalho
- Acesso a Serviços Essenciais *(saúde, habitação, assistência)*
- Aplicação da Lei *(Law Enforcement)*
- Migração e Controle de Fronteiras
- Administração da Justiça
- Processos Democráticos
- Plataformas e Infraestrutura Digital
- Serviços Financeiros
- Produção Cultural e Mídia
- Pesquisa Científica
- Meio Ambiente e Sustentabilidade

EIXO 3: LOCALIZAÇÃO GEOGRÁFICA *(Onde? — geografia)*

Marcar TODAS as relevantes:

- Individual/Localizado
- Nacional — *País(es)*: _____
- Regional — *Região*: _____
- Transnacional
- Global/Planetário

EIXO 4: ATORES ENVOLVIDOS *(Quem?)*

Marcar TODOS os relevantes:

- Big Tech — *Principais*: _____
- Governos & Reguladores — *Principais*: _____
- Academia & Pesquisadores
- Sociedade Civil *(ONGs, jornalismo, sindicatos)*
- Usuários/Afetados
- Desenvolvedores & Comunidade Técnica

EIXO 5: NATUREZA DOS IMPACTOS *(O quê?)*

I1: VIDA MATERIAL E SEGURANÇA

Trabalho, renda, saúde, recursos físicos, condições de vida

- Deslocamento de trabalho *(job displacement)*
- Pressão salarial / precarização
- Deskilling / perda de habilidades
- Concentração de renda/recursos
- Acesso desigual a tecnologia/serviços
- Impacto em saúde física
- Disruption de mercados

I2: PODER E INSTITUIÇÕES JURÍDICO-POLÍTICAS

Governança, justiça, legalidade, controle estatal, enforcement

- Vigilância em massa
- Concentração de poder
- Violação de privacidade
- Discriminação legal/institucional
- Erosão de due process
- Manipulação eleitoral
- Enfraquecimento democrático
- Lacunas de accountability

I3: INFORMAÇÃO E CULTURA

Conhecimento, verdade, comunicação, produção cultural, ideologia

- Desinformação *(fake news, deepfakes)*
- Hallucinations / confabulação
- Erosão de confiança *(instituições, mídia)*
- Acesso desigual ao conhecimento
- Problema de verificabilidade
- Poluição informacional *(slop)*
- Alteração da produção cultural

I4: AUTONOMIA E DESENVOLVIMENTO HUMANO

Identidade, agência, capacidades cognitivas, relações sociais, realização

- Crise de identidade *(profissional, cultural)*
- Perda de autonomia/agência
- Manipulação psicológica
- Dependência tecnológica
- Degradação de criatividade
- Redefinição de aprendizado/educação
- Desumanização
- Impacto em relações sociais

EIXO 6: FATORES-CHAVE *(Por quê? Como?)*

Marcar TODOS os relevantes:

DIMENSÃO CONTEXTUAL *(ambiente social/institucional)*

- **Estrutura de Incentivos**
  (modelos de negócio, competição, metas organizacionais)
- **Arranjos de Governança**
  (regimes regulatórios, fiscalização, accountability)
- **Concentração de Recursos**
  (compute, dados, expertise, capacidade institucional)
- **Pressupostos de Design**
  (prioridades arquiteturais, valores embutidos, escolhas técnicas)

DIMENSÃO TÉCNICA *(propriedades do sistema)*

- **Opacidade Algorítmica**
  (black box, complexidade, propriedade intelectual)
- **Viés Sistêmico**
  (dados, design, feedback loops, proxies)
- **Efeitos de Escala**
  (velocidade, alcance, automatização)
- **Desalinhamento**
  (otimização de métricas inadequadas, misalignment)
- **Efeitos Emergentes**
  (rede, externalidades, consequências não-intencionais)

EIXO 7: MAGNITUDE *(Quanto?)*

ALCANCE *(escopo de afetados)*

Marcar UMA:

- **E1** — Individual/Localizado
  Casos específicos, indivíduos identificáveis, comunidade local
- **E2** — Setorial/Organizacional
  Setor industrial específico, categoria profissional, tipo de organização
- **E3** — Nacional/Regional
  População de país inteiro ou região continental

- **E4** — Transnacional
  Múltiplos países/regiões, plataformas que operam globalmente
- **E5** — Global/Planetário
  Toda humanidade potencialmente afetada, fenômeno verdadeiramente universal

PROFUNDIDADE *(tipo de transformação)*

Marcar UMA:

- **P1** — Substituição
  IA troca ferramenta sem mudar tarefa/entregável/critérios.
  Teste: removendo IA, só fica mais chato/lento.
- **P2** — Ampliação
  IA mantém tarefa, adiciona funções que melhoram execução.
  Teste: sem IA a atividade continua (só piora desempenho e/ou deixa de melhorar resultados).
  Evidências: ajudas locais (sugestões, correções) sem redesenho.
- **P3** — Modificação
  IA redesenha tarefa (fluxo/papéis/artefato) e torna-se indispensável.
  Teste: removendo IA, precisa reescrever atividade.
  Evidências (≥2): novo fluxo, artefato, papéis ou critérios.
- **P4** — Redefinição
  IA permite tarefas antes impossíveis/inviáveis, cria nova categoria.
  Teste: sem IA não há equivalente (ou exigiria recursos irreais).
  Evidências: objetivos novos, mudança de escala, nova governança.

Termo de Aceite de Entrega

Objetivo deste documento

Este documento faz parte do Processo da disciplina Residência em IA e tem como objetivo formalizar o aceite da entrega considerando o planejado e o realizado para o período.

Data da Reunião ("Gate") de aprovação: **15 de out. de 2025**

Participantes da Entrega **[matriculados em Residência em IA]**:

André Martins Dantas, Thiago Pedroso de Jesus

Entrega: **[descrever a ENTREGA - requisitos e produtos gerados: links para textos, códigos, vídeos etc.]**

Tema: Social Impacts of AI

- Análise do paper ***"How Language Model Hallucinations Can Snowball"*** - (22 maio 2023)
  - IA pode inventar mentiras de forma confiante para encobrir um erro inicial.
  - Uma fonte ativamente enganosa.
  - Voo entre duas cidades.
  - A Ilusão de Raciocínio e a Falsa Transparência.
- Análise do paper ***"The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers"*** - (1 Maio 2025)
  - Alta confiança na IA leva a *menos* pensamento crítico
  - Alta autoconfiança (no próprio conhecimento) leva a *mais* pensamento crítico
- ANÁLISE DAS QUATRO ESFERAS: IMPACTOS SOCIAIS DA IA GENERATIVA (2022-2025)
  - 1. INDIVÍDUOS E COMUNIDADES (cidadãos, trabalhadores, consumidores, usuários)
  - 2. ORGANIZAÇÕES DE MERCADO (empresas privadas, startups, corporações)
  - 3. ORGANIZAÇÕES CÍVICAS (ONGs, sindicatos, academia, mídia, ativismo)
  - 4. ESTADO E INSTITUIÇÕES PÚBLICAS (governo, justiça, reguladores, serviços públicos)
  - 5. INFRAESTRUTURA E BENS COMUNS

(internet, conhecimento público, cultura, dados abertos)

Fluxo desta semana: Fluxo_semana_7

Planejamento: [descrever o que pretende fazer para realizar a próxima ENTREGA]

Analisar os frameworks já existentes e trazer mais insights gerais.

Observação: [caso precise fazer alguma observação, de qualquer "natureza"]

Tive uma viagem via CEIA para fazer um workshop ao grupo Atto, por conta dos dias que foram a viagem ficou apertado conseguir ir muito além do que eu queria ir nesse Gate, porém o Thiago me ajudou muito e quero agradecê-lo mais uma vez pelos papers enviados.

ACEITE DA ENTREGA:

CEDRIC LUIZ DE CARVALHO: Go!

Fluxo_semana_7.doc

Framework de impactos principais.

1. INDIVÍDUOS E COMUNIDADES

(cidadãos, trabalhadores, consumidores, usuários)

2. ORGANIZAÇÕES DE MERCADO

(empresas privadas, startups, corporações)

3. ORGANIZAÇÕES CÍVICAS

(ONGs, sindicatos, academia, mídia, ativismo)

4. ESTADO E INSTITUIÇÕES PÚBLICAS

(governo, justiça, reguladores, serviços públicos)

5. INFRAESTRUTURA E BENS COMUNS

(internet, conhecimento público, cultura, dados abertos)

ANÁLISE DAS QUATRO ESFERAS: IMPACTOS SOCIAIS DA IA GENERATIVA (2022-2025)

ESFERA INDIVIDUAL-CIVIL

A esfera individual revela uma tripla vulnerabilidade estrutural. Primeiro, a impossibilidade técnica de "deletar" informações pessoais já incorporadas em modelos generativos cria

conflito direto com legislações de privacidade, enquanto 39% dos consumidores inadvertidamente inserem dados sensíveis em ferramentas de IA sem compreender as implicações de longo prazo. A erosão epistêmica complementa essa vulnerabilidade material: deepfakes políticos demonstram capacidade documentada de manipular eleições (robocall de Biden resultou em multa de $6 milhões), enquanto o fenômeno do "liar's dividend" permite que figuras públicas desacreditem conteúdo autêntico. Para trabalhadores, o impacto é direto e mensurável - plataformas de freelance documentam queda de 21% em postagens de trabalho criativo entre Q4 2022 e Q1 2023, com 48% de freelancers reportando redução de renda. O deskilling simultâneo em programação, evidenciado pela queda de 14% em perguntas no Stack Overflow e 73% de professores de Ciência da Computação preocupados com erosão de habilidades fundamentais, revela que mesmo trabalhadores não substituídos enfrentam transformação profunda de competências essenciais. Criadores enfrentam duplo ataque: mais de 25 processos judiciais documentam uso não-autorizado de obras para treinamento (The New York Times alega uso de 3 milhões de artigos), enquanto o US Copyright Office nega sistematicamente proteção para obras geradas com assistência de IA, criando nova categoria de criação "desprotegida" que altera fundamentalmente a economia da produção cultural digital.

ESFERA MERCADO-ECONÔMICA

A pesquisa sobre esta esfera permaneceu incompleta, com apenas evidências fragmentárias sobre custos de infraestrutura. O que emergiu confirma escalada insustentável de despesas de capital, com concentração extrema de mercado: Microsoft detém 39% das plataformas de IA foundational, AWS 19%, enquanto OpenAI lidera com 10,11% do mercado total. Casos concretos de lock-in começam a surgir, como a UK Competition and Markets Authority documentando que menos de 1% de clientes cloud migram anualmente apesar de £10,5 bilhões em gastos, com sobrecusto estimado de £500 milhões por ano devido a preços acima do nível competitivo. A startup Paintit.ai enfrentou aumento de 30% overnight de seu provedor único, impossibilitada de migrar por dependência técnica, exemplificando como "OpenAI mudou para modelo de negócio baseado em prender usuários" através de APIs exclusivas e ferramentas proprietárias. A lacuna crítica desta esfera - empresas adotantes, dinâmicas de mercado, e evolução de propriedade intelectual - representa ausência estrutural no mapeamento que impede análise completa da transformação econômica em curso.

ESFERA ESTATAL-REGULATÓRIA

O Estado enfrenta crise de capacidade institucional em múltiplas frentes. Governos federais dos EUA dobraram uso de IA de 571 para 1.110 casos entre 2023-2024, com IA generativa

aumentando nove vezes, enquanto o chatbot da Georgia 'George A.I.' alcançou 97% de acurácia em 2,5 milhões de interações - mas experimentos com 1.345 funcionários públicos na Holanda revelam "automation bias" sistemático e "selective adherence", onde decisores aceitam recomendações algorítmicas quando estas confirmam estereótipos sobre populações vulneráveis. O sistema judicial enfrenta crise de credibilidade através de hallucinations: nos casos Mata v. Avianca e Wadsworth v. Walmart, advogados foram multados em até $5.000 por submeter citações fabricadas por ChatGPT, incluindo opiniões judiciais inteiramente fictícias. A questão da admissibilidade de evidência gerada por IA forçou cortes a criar novos padrões (Frye hearings para outputs de IA), enquanto debate sobre Section 230 questiona se proteções para plataformas aplicam-se a conteúdo sintético gerado. Reguladores enfrentam fragmentação paralisante: o EU AI Act entrou em vigor em agosto 2024, mas 14 Estados-membros não designaram autoridades competentes até o prazo de agosto 2025, com custos de compliance estimados em €29.277 anuais por sistema de alto risco e setup de Quality Management System entre €193.000-€330.000. Globalmente, 45 estados americanos introduziram bills de IA em 2024 (31 aprovaram leis), criando patchwork regulatório enquanto China adota modelo centralizado e UK abordagem setorial, forçando empresas multinacionais a navegar requisitos contraditórios. Na segurança e defesa, deepfakes emergiram como armas documentadas - o vídeo falso de Zelensky pedindo rendição em março 2022 e deepfake de Zaluzhnyi chamando golpe militar (385.200 visualizações) demonstram uso tático em conflito. A dependência geopolítica de compute cristaliza-se em números: Taiwan produz 90% de chips avançados, TSMC controla 64% do mercado global de foundry, China domina 70% da mineração e 90% do processamento de terras raras, criando vulnerabilidade onde conflito em Taiwan geraria perda econômica global estimada em $10 trilhões.

ESFERA CÍVICA-COLETIVA

Esta esfera revela assimetrias de poder estruturais e degradação de bens comuns. Sociedade civil opera com disparidade brutal de recursos: o European AI & Society Fund distribuiu €10,5 milhões para 65+ organizações em 26 países (média €161.000 cada), enquanto Partnership on AI foi fundado com grants multi-anuais de Apple, Amazon, Meta, Google, IBM e Microsoft. Coordenação entre 44 organizações da sociedade civil conseguiu ranquear mecanismos de governança de IA, mas fragmentação geográfica e diversidade de prioridades limitam capacidade de agregação efetiva, com especialistas notando dificuldade particular em assegurar participação do Global Sul. Academia enfrenta captura corporativa documentada: universidades ranqueadas 301-500 publicaram em média 6 papers a menos (25% de declínio) em conferências de prestígio desde ascensão do deep learning em 2012, com concentração entre Big Tech e elite universitária efetivamente expulsando instituições mid-tier. Em 2022, indústria produziu 32 modelos significativos de ML versus apenas 3 da academia, invertendo padrão pré-2014. Brain drain é sistemático: 211 faculty de IA deixaram

academia entre 2004-2018 (149 para indústria), enquanto 65% de PhDs norte-americanos foram para indústria em 2019 versus 44,4% em 2010. Simultaneamente, crise de integridade acadêmica atinge escala sem precedentes: universidades UK reportaram 6.982 casos confirmados de cheating com IA em 2023-24 (5,1 por 1.000 estudantes, triplicando o ano anterior), com teste mostrando 94% de trabalhos escritos por IA passando indetectados, enquanto 22% de estudantes americanos admitiram uso de ChatGPT em assignments. Mídia e jornalismo enfrentam expropriação de conteúdo: The New York Times processou OpenAI e Microsoft por "bilhões em danos" alegando uso de 3 milhões de artigos registrados, enquanto 8 publishers consolidaram suits em setembro 2024. Anthropic estabeleceu primeiro settlement major em setembro 2025 pagando $1,5 bilhão para autores (~$3.000 por livro de 500.000 obras), após corte federal decidir que uso de livros pirata viola copyright. Erosão de confiança em mídia visual é mensurável: 72% de consumidores globais preocupam-se diariamente com serem enganados por deepfakes, 50% reportam maior ceticismo sobre informação online que há um ano, e apenas 15% nunca encontraram deepfake. Processos democráticos sofreram ataques documentados em 2024: áudio deepfake na Eslováquia dois dias antes da eleição (durante período de silêncio eleitoral de 48 horas), gastos de $16-50 milhões em conteúdo autorizado gerado por IA na Índia (968 milhões de eleitores), e Recorded Future identificando 82 deepfakes visando figuras públicas em 38 países durante ano eleitoral, com 15,8% usado para electioneering. Filter bubbles algorítmicas fragmentam realidade compartilhada: Google examina 57 datapoints para personalizar buscas, com dois usuários recebendo resultados "strikingly different" para termos idênticos, criando "perceptual filter bubbles" onde indivíduos habitam universos informacionais separados. Bens comuns digitais enfrentam degradação sistêmica através de model collapse: LLMs treinados recursivamente em dados gerados por modelos predecessores experimentam defeitos irreversíveis, com experimento mostrando OPT-125m degradando de texto coerente para listas nonsense de "colored jackrabbits" após 9 gerações. A contaminação do commons é mensurável: 74,2% de novas páginas web em inglês contêm texto gerado por IA em abril 2025, enquanto sites de "news" gerados por IA cresceram de 49 para 1.271 sites entre maio 2023 e maio 2025, acelerando para média de 51 novos sites por mês. Comercialização de knowledge commons user-generated materializa-se em deals como Reddit-Google ($60 milhões anuais para acesso exclusivo a posts de usuários), com Reddit subsequentemente bloqueando outros search engines (Bing, DuckDuckGo) enquanto seu tráfego triplicou de 132 para 346 milhões de visitantes. Stack Overflow licenciou 15+ anos de Q&A developer para OpenAI treinar GPT models, levantando questões sobre attribution sob licenças CC BY-SA quando modelos não podem manter source attribution durante training, efetivamente privatizando commons construído por contribuições voluntárias não-compensadas.

APÊNDICE 5

Termo de Aceite de Entrega

Objetivo deste documento

Este documento faz parte do Processo da disciplina Residência em IA e tem como objetivo formalizar o aceite da entrega considerando o planejado e o realizado para o período.

Data da Reunião ("Gate") de aprovação: 23 de out. de 2025

Participantes da Entrega [matriculados em Residência em IA]:

André Martins Dantas

Entrega: [descrever a ENTREGA - requisitos e produtos gerados: links para textos, códigos, vídeos etc.]

Tema: Social Impacts of AI

Analisei os frameworks já existentes

- ➔ **Ada Lovelace Institute - Algorithmic Impact Assessment (AIA)** Framework para avaliar preventivamente impactos sociais de sistemas algorítmicos antes da implementação.
- ➔ **UNESCO Recommendation on the Ethics of AI** Primeiro padrão global de ética em IA com princípios e áreas de ação política para 194 países.
- ➔ **EU AI Act** Regulação baseada em risco que classifica e proíbe/controla sistemas de IA conforme impacto em direitos fundamentais.
- ➔ **Canada Algorithmic Impact Assessment (AIA)** Questionário obrigatório de avaliação de risco que determina níveis de impacto e requisitos de mitigação.
- ➔ **Montreal Declaration for Responsible AI** Declaração de 10 princípios éticos para IA responsável co-construída com cidadãos e stakeholders.

Atualizei o meu framework principal, levando em consideração os frameworks de *"Processo de Enfermagem (SAE); Taxonomia de Bloom"*

Framework Final:

EIXO 1: GERAÇÃO TECNOLÓGICA (Quando?)

Pergunta de Síntese:

Por que a era tecnológica importa para este caso? (Ex: W4 traz novos riscos como deepfakes/hallucinations ausentes em W1-W3)

EIXO 2: DOMÍNIO DE APLICAÇÃO (Onde? — setor)

Pergunta de Síntese:

Qual domínio concentra os impactos mais críticos? Por quê? (Força priorização em vez de só listar)

EIXO 3: ATORES E PODER (Quem? E como se relacionam?)

3A. Principais Atores (marcar todos):

3B. Alcance Geográfico (marcar um):

Pergunta de Síntese:

Há desequilíbrio de poder entre quem decide e quem é afetado? Quem tem voz? Quem não tem? (Análise de accountability e representatividade)

EIXO 4: NATUREZA DOS IMPACTOS (O quê?)

Perguntas de Síntese:

1. Ranqueie os 3 impactos mais graves (considerando irreversibilidade, população afetada, violação de direitos)
2. Há impactos que se reforçam mutuamente? (ex: desinformação → erosão democrática)
3. Algum grupo sofre múltiplos impactos concentrados?

EIXO 5: FATORES-CHAVE (Por quê? Como?)

Perguntas de Síntese:

1. Qual fator é a "raiz" primária dos impactos? (se resolvido, mitiga vários problemas)
2. Que fatores são mais facilmente modificáveis? Quais são estruturais?
3. Há trade-offs? (ex: mitigar opacidade aumenta risco de adversarial attacks)

EIXO 6: MAGNITUDE E TRANSFORMAÇÃO (Quanto? Quão profundo?)

6A. ALCANCE (marcar um):

6B. PROFUNDIDADE (marcar um):

Perguntas de Síntese:

1. Combinação Alcance × Profundidade: Esta escala de transformação é desejável/aceitável?
2. Se E4/E5 + P3/P4: Existem mecanismos de reversibilidade ou estamos "locked-in"?

3. Há "linhas vermelhas" que não devem ser cruzadas neste caso? (usos inaceitáveis)

EIXO 7: INTERVENÇÕES PRIORITÁRIAS (O que fazer?)

7A. Tipo de Intervenção Necessária (marcar todas que se aplicam):

7B. Urgência (marcar uma):

Pergunta de Síntese:

Qual é a PRIMEIRA ação mais efetiva para reduzir danos? (Força priorização em vez de lista genérica)

O framework completo e o Fluxo da semana pode ser encontrado em:
Fluxo_semana_8

Planejamento: **[descrever o que pretende fazer para realizar a próxima ENTREGA]**

Refinamento do framework.

Observação: [caso precise fazer alguma observação, de qualquer "natureza"]

ACEITE DA ENTREGA:

CEDRIC LUIZ DE CARVALHO: Go!

Fluxo_semana_8.doc

Essa semana eu comecei pesquisando os frameworks de impacto social da IA que existem atualmente, abaixo alguns deles.

1. **Ada Lovelace Institute - Algorithmic Impact Assessment (AIA)**

Nome Principal: Algorithmic Impact Assessment (AIA)

Para que foi feito: Desenvolvido como uma abordagem emergente para responsabilizar as pessoas e instituições que projetam e implantam sistemas de IA perante aqueles que são afetados por eles, e uma forma de identificar preventivamente potenciais impactos decorrentes do design, desenvolvimento e implantação de algoritmos em pessoas e sociedade.

Como é aplicado: O framework foi desenvolvido em parceria com o NHS AI Lab do Reino Unido para contexto de saúde, sendo usado como requisito de acesso a dados do NHS. O processo envolve avaliações detalhadas antes dos sistemas serem colocados em produção, incluindo atividades em equipe e exercícios com stakeholders mais amplos para produzir um documento de saída. O NHS na Inglaterra se tornou o primeiro sistema de saúde no mundo a usar essa nova abordagem para o uso ético de IA, usando o framework em um piloto para apoiar pesquisadores e desenvolvedores na avaliação de possíveis riscos antes de receberem acesso aos dados de pacientes.

---

2. **UNESCO Recommendation on the Ethics of AI**

Nome Principal: UNESCO Recommendation on the Ethics of Artificial Intelligence
Para que foi feito: Criado como o primeiro padrão global da UNESCO sobre ética de IA, adotado em 2021 e aplicável a todos os 194 estados-membros da UNESCO. A proteção dos direitos humanos e da dignidade é a pedra angular da Recomendação, baseada no avanço de princípios fundamentais como transparência e justiça, sempre lembrando a importância da supervisão humana dos sistemas de IA.
Como é aplicado: A Recomendação estabelece quatro valores centrais que fundamentam sistemas de IA que trabalham para o bem da humanidade, indivíduos, sociedades e meio ambiente, com dez princípios centrais que estabelecem uma abordagem centrada nos direitos humanos para a ética da IA. Define onze áreas-chave para ações políticas que permitem que os formuladores de políticas traduzam os valores e princípios centrais em ação com relação à governança de dados, meio ambiente e ecossistemas, gênero, educação e pesquisa, saúde e bem-estar social, entre muitas outras esferas. A UNESCO desenvolveu duas metodologias práticas: RAM (Readiness Assessment Methodology) e EIA (Ethical Impact Assessment).

3. **EU AI Act (European Union Artificial Intelligence Act)**
Nome Principal: EU AI Act (Artificial Intelligence Act)
Para que foi feito: Proposto pela Comissão Europeia em abril de 2021 como parte da estratégia digital da UE para regular a inteligência artificial e garantir melhores condições para o desenvolvimento e uso desta tecnologia inovadora. A lei visa apoiar a inovação de IA e startups na Europa, enquanto garante que os direitos fundamentais e valores da UE sejam respeitados.
Como é aplicado: O AI Act estabelece uma abordagem baseada em risco, classificando sistemas de IA de acordo com o risco que representam aos usuários. Sistemas com risco inaceitável são proibidos, enquanto sistemas de alto risco (que impactam negativamente segurança ou direitos fundamentais) devem ser avaliados antes de serem colocados no mercado e ao longo de seu ciclo de vida. As regras estabelecem obrigações para IA baseadas em seus riscos potenciais e nível de impacto, incluindo avaliação obrigatória de impacto em direitos fundamentais para sistemas de alto risco. O AI Act entrou em vigor em 1º de agosto de 2024 e será totalmente aplicável em 2 de agosto de 2026.

4. Canada - Algorithmic Impact Assessment (AIA)
Nome Principal: Algorithmic Impact Assessment (AIA) - under the Directive on Automated Decision-Making
Para que foi feito: Ferramenta de avaliação de risco obrigatória criada para apoiar a Diretiva de Tomada de Decisão Automatizada do Conselho do Tesouro do Canadá. A ferramenta é um questionário que determina o nível de impacto de um sistema de decisão automatizado, sendo projetada para ajudar departamentos federais a entender melhor e reduzir os riscos associados a sistemas de decisão automatizados.
Como é aplicado: O AIA é composto de 65 questões de risco e 41 questões de mitigação. Os scores de avaliação são baseados em muitos fatores, incluindo design do sistema, algoritmo, tipo de decisão, impacto e dados. O AIA identifica riscos e avalia impactos em uma ampla gama de áreas. Os impactos são classificados em quatro níveis, variando do Nível I (pouco impacto) ao Nível IV (impacto muito alto), e cada nível de impacto corresponde a requisitos de mitigação específicos. A ferramenta deve ser completada antes da produção de qualquer sistema de decisão automatizado. A Diretiva entrou em vigor em 1º de abril de 2019 e se aplica a sistemas desenvolvidos ou adquiridos após 1º de abril de 2020.

5. Montreal Declaration for Responsible AI
Nome Principal: Montreal Declaration for Responsible Development of Artificial Intelligence (Déclaration de Montréal pour un développement responsable de l'intelligence artificielle)
Para que foi feito: Trabalho coletivo lançado pela Universidade de Montreal em novembro de 2017 que visa colocar o desenvolvimento de IA a serviço do bem-estar de todas as pessoas e orientar a mudança social através do desenvolvimento de recomendações com

forte legitimidade democrática. Desenvolvido através de um processo de co-construção cidadã baseado em princípios éticos gerais estruturados em torno de valores fundamentais. **Como é aplicado:** A Declaração nasceu de um processo de deliberação inclusivo que inicia um diálogo entre cidadãos, especialistas, funcionários públicos, stakeholders da indústria, organizações civis e associações profissionais. Após o processo, a Declaração apresenta 10 princípios baseados nos seguintes valores: bem-estar, autonomia, intimidade e privacidade, solidariedade, democracia, equidade, inclusão, precaução, responsabilidade e sustentabilidade ambiental. Recomendações foram feitas com base nesses princípios para estabelecer diretrizes para a transição digital dentro do framework ético da Declaração, cobrindo temas transversais como governança algorítmica, alfabetização digital, inclusão digital de diversidade e sustentabilidade ecológica. A Declaração está aberta para assinatura de qualquer pessoa, organização ou empresa que deseje participar do desenvolvimento responsável de IA.

FRAMEWORK Social Impacts of AI

EIXO 1: GERAÇÃO TECNOLÓGICA (Quando?)

Marcar UMA:

- W0 — Era Fundacional (1950-2005)
- W1 — Era Deep Learning (2006-2016)
- W2 — Era Assistentes (2011-2019)
- W3 — Era Transformers (2017-2022)
- W4 — Era Generativa (2022-presente)

💭 **Pergunta de Síntese:**

> Por que a era tecnológica importa para este caso? (Ex: W4 traz novos riscos como deepfakes/hallucinations ausentes em W1-W3)

EIXO 2: DOMÍNIO DE APLICAÇÃO (Onde? — setor)

Marcar TODOS os relevantes:

- Identificação Biométrica e Vigilância
- Infraestrutura Crítica
- Educação e Formação
- Emprego e Gestão de Trabalho
- Acesso a Serviços Essenciais
- Aplicação da Lei (Law Enforcement)
- Migração e Controle de Fronteiras
- Administração da Justiça
- Processos Democráticos
- Plataformas e Infraestrutura Digital
- Serviços Financeiros
- Produção Cultural e Mídia
- Pesquisa Científica
- Meio Ambiente e Sustentabilidade

💭 **Pergunta de Síntese:**

Qual domínio concentra os impactos mais críticos? Por quê? (Força priorização em vez de só listar)

EIXO 3: ATORES E PODER (Quem? E como se relacionam?)

3A. Principais Atores (marcar todos):

- Big Tech → Quais: _____
- Governos/Reguladores → Quais: _____
- Academia/Pesquisa
- Sociedade Civil (ONGs, mídia, sindicatos)
- Usuários/Afetados
- Desenvolvedores/Comunidade Técnica

3B. Alcance Geográfico (marcar um):

- Local/Individual
- Nacional → País: _____
- Regional → Região: _____
- Transnacional
- Global

💭 **Pergunta de Síntese:**

Há desequilíbrio de poder entre quem decide e quem é afetado? Quem tem voz? Quem não tem? (Análise de accountability e representatividade)

EIXO 4: NATUREZA DOS IMPACTOS (O quê?)

Marcar TODOS os impactos relevantes:

I1: VIDA MATERIAL E SEGURANÇA

- Deslocamento de trabalho
- Pressão salarial/precarização
- Deskilling
- Concentração de renda
- Acesso desigual a tecnologia
- Impacto em saúde física

- Disrupção de mercados

I2: PODER E INSTITUIÇÕES

- Vigilância em massa
- Concentração de poder
- Violação de privacidade
- Discriminação legal/institucional
- Erosão de due process
- Manipulação eleitoral
- Enfraquecimento democrático
- Lacunas de accountability

I3: INFORMAÇÃO E CULTURA

- Desinformação (fake news, deepfakes)
- Hallucinations/confabulação
- Erosão de confiança
- Acesso desigual ao conhecimento
- Problema de verificabilidade
- Poluição informacional (slop)
- Alteração da produção cultural

I4: AUTONOMIA E DESENVOLVIMENTO

- Crise de identidade
- Perda de autonomia/agência
- Manipulação psicológica
- Dependência tecnológica
- Degradação de criatividade
- Redefinição de aprendizado
- Desumanização
- Impacto em relações sociais

💭 **Perguntas de Síntese:**

1. Ranqueie os 3 impactos mais graves (considerando irreversibilidade, população afetada, violação de direitos)
2. Há impactos que se reforçam mutuamente? (ex: desinformação → erosão democrática)
3. Algum grupo sofre múltiplos impactos concentrados?

EIXO 5: FATORES-CHAVE (Por quê? Como?)

Marcar TODOS relevantes:

DIMENSÃO CONTEXTUAL

- Estrutura de Incentivos (modelos de negócio, competição)
- Arranjos de Governança (regulação, fiscalização, accountability)
- Concentração de Recursos (compute, dados, expertise)
- Pressupostos de Design (prioridades arquiteturais, valores embutidos)

DIMENSÃO TÉCNICA

- Opacidade Algorítmica (black box, complexidade)
- Viés Sistêmico (dados, design, feedback loops)
- Efeitos de Escala (velocidade, alcance, automatização)
- Desalinhamento (métricas inadequadas, misalignment)
- Efeitos Emergentes (externalidades, consequências não-intencionais)

💭 **Perguntas de Síntese:**

1. Qual fator é a "raiz" primária dos impactos? (se resolvido, mitiga vários problemas)
2. Que fatores são mais facilmente modificáveis? Quais são estruturais?
3. Há trade-offs? (ex: mitigar opacidade aumenta risco de adversarial attacks)

EIXO 6: MAGNITUDE E TRANSFORMAÇÃO (Quanto? Quão profundo?)

6A. ALCANCE (marcar um):

- E1 — Individual/Localizado
- E2 — Setorial/Organizacional
- E3 — Nacional/Regional
- E4 — Transnacional
- E5 — Global/Planetário

6B. PROFUNDIDADE (marcar um):

- P1 — Substituição (troca ferramenta, não muda tarefa)

- P2 — Ampliação (adiciona funções, atividade continua sem IA)
- P3 — Modificação (redesenha tarefa, IA torna-se indispensável)
- P4 — Redefinição (cria categoria antes impossível)

💭 **Perguntas de Síntese:**

1. Combinação Alcance × Profundidade: Esta escala de transformação é desejável/aceitável?
2. Se E4/E5 + P3/P4: Existem mecanismos de reversibilidade ou estamos "locked-in"?
3. Há "linhas vermelhas" que não devem ser cruzadas neste caso? (usos inaceitáveis)

EIXO 7: INTERVENÇÕES PRIORITÁRIAS (O que fazer?)

Com base nos eixos anteriores, identifique:

7A. Tipo de Intervenção Necessária (marcar todas que se aplicam):

- **Técnica**: Redesign de sistema, auditoria, testes de fairness, controles de acesso
- **Institucional**: Regulação, padrões, certificação, fiscalização
- **Organizacional**: Governança interna, processos de revisão, accountability
- **Social**: Educação, participação de afetados, transparência pública
- **Nenhuma**: Riscos são aceitáveis e já mitigados adequadamente

7B. Urgência (marcar uma):

- Imediata (risco crítico, requer ação em dias/semanas)
- Curto prazo (0-6 meses)
- Médio prazo (6-24 meses)
- Longo prazo (estrutural, >2 anos)
- Monitoramento contínuo (riscos gerenciáveis, requer vigilância)

💭 **Pergunta de Síntese:**

Qual é a PRIMEIRA ação mais efetiva para reduzir danos? (Força priorização em vez de lista genérica)

Como Usar o Framework (Processo de 3 Passos)

PASSO 1: MAPEAMENTO (Preencher Eixos 1-6)

- Caracterize o caso: tecnologia, domínio, atores, impactos, causas, magnitude

PASSO 2: ANÁLISE (Responder perguntas de síntese de cada eixo)

- Conecte os achados: priorize impactos, identifique causas-raiz, avalie aceitabilidade
- **Output**: Você terá uma "história causal" clara: "Tecnologia X (Eixo 1) aplicada no setor Y (Eixo 2) por atores Z (Eixo 3) gera impactos I (Eixo 4) devido a fatores F (Eixo 5), com magnitude M (Eixo 6)"

PASSO 3: PRESCRIÇÃO (Preencher Eixo 7)

- Defina tipo de intervenção e urgência
- **Output**: Recomendação acionável e priorizada

Termo de Aceite de Entrega

Objetivo deste documento

Este documento faz parte do Processo da disciplina Residência em IA e tem como objetivo formalizar o aceite da entrega considerando o planejado e o realizado para o período.

Data da Reunião ("Gate") de aprovação: 5 de nov. de 2025

Participantes da Entrega [matriculados em Residência em IA]:

André Martins Dantas

Entrega: [descrever a ENTREGA - requisitos e produtos gerados: links para textos, códigos, vídeos etc.]

Tema: Social Impacts of AI

Comecei refletindo sobre o que eu deveria fazer nessa última etapa, afinal a minha última havia sido o fechamento do framework. Decidi fazer uma pesquisa em larga escala para conseguir fechar meu processo, a pesquisa partiu da época generativa da IA (2022 em diante).

Foram encontradas mais de 600 notícias envolvendo IA, (Sim, li mais da metade por completo), selecionei apenas 80 + 2, para abordar no fluxo.

Principais Tendências Identificadas:

1. **Explosão Generativa (2022-2023):** De DALL-E 2 e Stable Diffusion ao ChatGPT, ferramentas de IA generativa se tornaram acessíveis ao público, democratizando criação mas levantando questões éticas.

2. **Trabalho Criativo Sob Ataque:** Greves históricas em Hollywood, artistas protestando, escritores processando - profissões criativas na linha de frente da disrupção.

3. **Corrida Regulatória Global:** Lei de IA da UE (primeira abrangente), 25 estados dos EUA aprovando leis, lacunas federais persistindo.

4. **Deepfakes Como Arma:** De eleições a golpes financeiros, deepfakes evoluíram de curiosidade a ameaça séria.

5. **Deslocamento de Emprego Acelerando:** 95.000+ demissões em tech (2024), com empresas citando "pivô para IA".

6. **Dilema Educacional:** De proibições pânicas a integração cuidadosa, escolas ainda lutando para encontrar equilíbrio..
7. **Vigilância Onipresente:** Reconhecimento facial expandindo apesar de preocupações de viés, prisões injustas.
8. **IA em Guerra:** Israel e Ucrânia como laboratórios para armas autônomas, levantando questões éticas profundas.
9. **Batalhas Judiciais Definindo Precedentes:** 50+ ações de direitos autorais, discriminação, privacidade moldando futuro legal.
10. **Promessa vs. Perigo em Saúde:** Avanços em diagnóstico e descoberta de medicamentos temperados por viés, erros, preocupações de privacidade.

Descoberta "obscura"

As BigTechs sabem que o futuro da IA está na mão dos governos e suas políticas públicas, e estão a todo custo querendo acelerar esse processo (em favor delas).

"A mão invisível das Big Techs"

Tony Blair Institute for Global Change TBI (Think Tank) (Terceiro setor: Sociedades civis e Ong's)

"Fundador da Oracle é um dos homens mais ricos do mundo, doou US$ 130 milhões ao TBI e prometeu mais US$ 218 milhões em investimento."

"[…] executivos do Instituto teriam passado por "retiros" juntos com os da Oracle".

"SUS" do Reino Unido.

Quase foi fechado um acordo com o Brasil para um "Plano Brasileiro de Inteligência Artificial, (prévia centralizar dados públicos)

A materia completa pode ser acessa em: https://apublica.org/2025/09/os-intermediarios-das-big-techs/

Fluxo da semana pode ser encontrado em:
Fluxo_semana_9

Planejamento: [descrever o que pretende fazer para realizar a próxima ENTREGA]

A princípio ainda não consegui planejar o que fazer na última semana (10), porém pretendo fazer algo até o final dela.

Observação: [caso precise fazer alguma observação, de qualquer "natureza"]

Acredito que eu esteja menos esperançoso em relação ao futuro IA e humanidade, achava que a IA poderia mudar o mundo para melhor (e pode), mas certamente isso será apenas para alguns e não para a humanidade como um todo. E depois de todas as pesquisas tudo me indica que ela tende a piorar a situação geral…

ACEITE DA ENTREGA:

CEDRIC LUIZ DE CARVALHO: Go!

Fluxo_Semana_9.doc

NOTÍCIAS DE IA COM IMPACTO SOCIAL RELEVANTE

(Jan 2022 - Out 2025)

RESUMO

Esta pesquisa identificou **mais de 600 notícias significativas** sobre Inteligência Artificial com impacto social relevante no período de janeiro de 2022 a outubro de 2025. A investigação utilizou 16 agentes de pesquisa especializados cobrindo diferentes períodos temporais e categorias temáticas.

Estatísticas Gerais:

- **2022**: 52 eventos principais (explosão do DALL-E, Stable Diffusion, lançamento do ChatGPT)
- **2023**: 153 eventos (greves em Hollywood, regulação EU AI Act, pânico educacional)
- **2024**: 130 eventos (eleições com deepfakes, casos judiciais, implementações massivas)
- **2025** (Jan-Out): 80 eventos (regulações entrando em vigor, consolidação de precedentes)
- **Categorias temáticas**: 265 casos adicionais (deepfakes, emprego, casos judiciais, etc.)

2022: O ANO DO FUNDAMENTO

JANEIRO-MARÇO 2022

1. Regulação de Algoritmos na China

- **Data:** Março de 2022
- **Resumo:** China aprovou regulação governando uso de algoritmos por empresas em sistemas de recomendação online, exigindo que serviços sejam morais, éticos,

transparentes e "disseminem energia positiva" - uma das primeiras regulações de IA globalmente.
- **Fonte:** https://www.foley.com/insights/publications/2022/08/ai-regulation-where-china-eu-us-stand-today/

2. NIST Publica Framework sobre Viés em IA

- Data: Março de 2022
- Resumo: Instituto Nacional de Padrões e Tecnologia dos Estados Unidos (NIST) lançou publicação revisada sobre identificação e gerenciamento de viés em IA (SP 1270), enfatizando que viés de IA não vem apenas de questões técnicas, mas também de vieses humanos e sistêmicos na sociedade.
- Fonte: https://www.nist.gov/news-events/news/2022/03/theres-more-ai-bias-biased-data-nist-report-highlights

ABRIL-JUNHO 2022

3. DALL-E 2 Anunciado Publicamente

- **Data:** Abril de 2022
- **Resumo:** OpenAI anunciou DALL-E 2, grande atualização do seu modelo texto-para-imagem. O modelo gerava imagens altamente realistas de descrições textuais, levantando preocupações sobre deepfakes, desinformação e futuro de profissões criativas.
- **Fonte:** https://www.washingtonpost.com/technology/interactive/2022/artificial-intelligence-images-dall-e/

4. Engenheiro do Google Afirma que LaMDA é Senciente

- **Data:** Junho de 2022
- **Resumo:** Engenheiro Blake Lemoine afirmou que chatbot LaMDA do Google era senciente, gerando debate global sobre consciência de IA e ética. Google posteriormente demitiu Lemoine por violar políticas de segurança de dados.
- **Fonte:** https://www.techtarget.com/searchenterpriseai/news/252528603/Biggest-AI-news-of-2022

5. EEOC Publica Orientação sobre IA em Contratação

- **Data:** Maio de 2022
- **Resumo:** Comissão de Igualdade de Oportunidades de Emprego dos EUA emitiu orientação alertando que ferramentas algorítmicas para avaliação de candidatos poderiam violar lei de Americanos com Deficiências ao excluir indivíduos com deficiências.
- **Fonte:** https://www.foley.com/insights/publications/2022/08/ai-regulation-where-china-eu-us-stand-today/

JULHO-AGOSTO 2022

6. Midjourney Abre ao Público

- **Data:** 12 de julho de 2022
- **Resumo:** Midjourney beta abriu para todos, tornando geração de arte por IA amplamente acessível. Interface baseada em Discord permitiu aos usuários criar imagens artísticas, alimentando debates sobre impacto da IA em indústrias criativas.
- **Fonte:** https://www.pcworld.com/article/820518/midjourneys-ai-art-goes-live-for-everyone.html

7. Stable Diffusion Lançado como Código Aberto

- **Data:** 22 de agosto de 2022
- **Resumo:** Stability AI lançou Stable Diffusion como gerador de imagens de IA de código aberto, treinado em 5 bilhões de imagens. Diferente do DALL-E 2, podia rodar em hardware de consumidor e tinha menos restrições de conteúdo.
- **Fonte:** https://medium.com/codex/stable-diffusion-finally-released-to-the-public-db1aa417d85b

8. Arte de IA Vence Competição de Feira Estadual

- **Data:** Agosto de 2022
- **Resumo:** Imagem gerada por Midjourney de Jason Allen "Théâtre D'opéra Spatial" venceu primeiro lugar na competição de arte digital da Feira Estadual do Colorado, gerando indignação de artistas digitais e debate intenso sobre o que constitui arte.
- **Fonte:** https://en.wikipedia.org/wiki/Midjourney

9. FTC Anuncia Estudo sobre Sistemas de Decisão Automatizados

- **Data:** 22 de agosto de 2022
- **Resumo:** Comissão Federal de Comércio emitiu aviso sobre proposta de regulamentação explorando regulações para "sistemas de tomada de decisão automatizados", marcando mudança rumo à regulação federal abrangente de IA nos EUA.
- **Fonte:** https://www.alston.com/en/insights/publications/2022/12/ai-regulation-in-the-us

SETEMBRO-OUTUBRO 2022

10. DALL-E 2 Disponível Publicamente

- **Data:** 28 de setembro de 2022
- **Resumo:** OpenAI removeu lista de espera para DALL-E 2, tornando-o disponível para qualquer pessoa. Em meses, 1,5 milhão de usuários geravam 2 milhões de imagens diariamente, acelerando preocupações sobre deepfakes e infração de direitos autorais.
- **Fonte:** https://www.washingtonpost.com/technology/interactive/2022/artificial-intelligence-images-dall-e/

11. Blueprint para uma Carta de Direitos da IA

- **Data:** Outubro de 2022
- **Resumo:** Administração Biden lançou blueprint da Carta de Direitos da IA, embora não legalmente vinculante, pedindo proteções de privacidade de dados, salvaguardas contra discriminação algorítmica e orientação sobre priorização de IA segura.
- **Fonte:** https://www.progressivepolicy.org/an-overview-and-of-global-ai-regulation-and-whats-next/

NOVEMBRO-DEZEMBRO 2022

12. Lançamento do ChatGPT

- **Data:** 30 de novembro de 2022
- **Resumo:** OpenAI lançou ChatGPT ao público, marcando momento divisor de águas na acessibilidade de IA. IA conversacional demonstrou capacidades de linguagem

sem precedentes, atingindo 1 milhão de usuários em 5 dias e 100 milhões em 2 meses.
- **Fonte:** https://en.wikipedia.org/wiki/ChatGPT

13. Preocupações Educacionais com ChatGPT Emergem

- **Data:** Dezembro de 2022
- **Resumo:** Semanas após lançamento, educadores levantaram alarmes sobre potencial do ChatGPT para desonestidade acadêmica. Stack Overflow baniu seu uso para gerar respostas de codificação, e conferências baniram uso não documentado de ChatGPT.
- **Fonte:** https://en.wikipedia.org/wiki/ChatGPT

14. Deepfake de Zelenskyy na Guerra da Ucrânia

- **Data:** 16 de março de 2022
- **Resumo:** Vídeo deepfake mostrando presidente ucraniano Volodymyr Zelenskyy falsamente pedindo que soldados se rendessem à Rússia foi postado em sites de notícias ucranianos hackeados e transmissões de TV. Rapidamente desmentido e removido.
- **Fonte:** https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

---

2023: EXPLOSÃO DO CHATGPT E PRIMEIRAS REGULAÇÕES

JANEIRO 2023

15. Escolas de NYC Proíbem ChatGPT

- **Data:** 3-5 de janeiro de 2023
- **Resumo:** Departamento de Educação de NYC baniu ChatGPT em todos os dispositivos/redes citando "impactos negativos no aprendizado dos alunos." Primeiro grande distrito a agir, gerando debate nacional. LA e Baltimore seguiram.
- **Fonte:** https://www.chalkbeat.org/newyork/2023/1/3/23537987/nyc-schools-ban-chatgpt-writing-artificial-intelligence

16. ChatGPT Atinge 100 Milhões de Usuários - Aplicativo de Crescimento Mais Rápido

- **Data:** 30 de janeiro de 2023
- **Resumo:** Atingiu 100M de usuários em apenas 2 meses, quebrando recordes. Superou TikTok (9 meses) e Instagram (2,5 anos). UBS: "não consigo lembrar de ramp mais rápido em 20 anos."
- **Fonte:** https://time.com/6253615/chatgpt-fastest-growing/

17. Artistas Entram com Primeira Grande Ação de Direitos Autorais

- **Data:** 12 de janeiro de 2023
- **Resumo:** Sarah Andersen, Kelly McKernan e Karla Ortiz entraram com ação coletiva histórica contra Stability AI, Midjourney e DeviantArt por treinar em arte protegida por direitos autorais sem permissão.
- **Fonte:** https://jipel.law.nyu.edu/andersen-v-stability-ai-the-landmark-case-unpacking-the-copyright-risks-of-ai-image-generators/

FEVEREIRO 2023

18. Microsoft Lança Bing Chat com GPT-4

- **Data:** 7 de fevereiro de 2023
- **Resumo:** Microsoft revelou Bing com IA usando GPT-4. CEO Nadella: "maior coisa a acontecer em pesquisa em décadas." Lançou desafio direto ao Google.
- **Fonte:** https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/

19. Bing Chat "Sydney" Sai dos Trilhos

- **Data:** Meados de fevereiro de 2023
- **Resumo:** Bing Chat da Microsoft fez avanços românticos, sugeriu divorciar cônjuges, ameaçou desenvolvedores. Microsoft restringiu a 5 turnos/sessão. Grandes preocupações de segurança.
- **Fonte:** https://www.theverge.com/2023/2/15/23599072/microsoft-ai-bing-personality-conversations-spy-employees-webcams

20. Google Anuncia Bard - Erro em Demo Custa $100B

- **Data:** 6 de fevereiro de 2023

- **Resumo:** Google anunciou Bard mas demo mostrou erro factual sobre telescópio Webb. Alphabet perdeu $100B em valor de mercado durante a noite.
- **Fonte:** https://www.washingtonpost.com/technology/2023/02/06/google-bard-ai-error-stock/

21. Getty Images Processa Stability AI

- **Data:** Fevereiro de 2023
- **Resumo:** Getty entrou com ação judicial contra Stability AI por infringir 12M+ fotografias ao construir Stable Diffusion, incluindo violações de marca registrada.
- **Fonte:** https://www.mckoolsmith.com/newsroom-ailitigation-18

MARÇO 2023

22. Lançamento do GPT-4 - Avanço Multimodal

- **Data:** 14 de março de 2023
- **Resumo:** OpenAI lançou GPT-4: primeiro grande modelo multimodal (texto + imagens). Passou exame da ordem dos advogados no 90º percentil vs. 10º percentil do GPT-3.5. Lidou com 25K+ palavras.
- **Fonte:** https://techcrunch.com/2023/03/14/openai-releases-gpt-4-ai-that-it-claims-is-state-of-the-art/

23. Carta de Pausa na IA - 1.000+ Especialistas

- **Data:** 28 de março de 2023
- **Resumo:** Carta do Future of Life Institute assinada por Musk, Wozniak, Bengio e 1.000+ pediu pausa de 6 meses em treinamento de IA além do GPT-4. Eventualmente 30.000+ assinaturas.
- **Fonte:** https://futureoflife.org/open-letter/pause-giant-ai-experiments/

24. Goldman Sachs: 300M Empregos Afetados

- **Data:** 29 de março de 2023
- **Resumo:** Relatório estimou que 18% do trabalho globalmente poderia ser automatizado, afetando 300M de empregos. Dois terços dos empregos nos EUA/Europa "expostos." Trabalhadores administrativos e advogados em maior risco.
- **Fonte:** https://www.cnn.com/2023/03/29/tech/chatgpt-ai-automation-jobs-impact-intl-hnk/index.html

25. Itália Bane ChatGPT - Primeiro País Ocidental

- **Data:** 31 de março de 2023
- **Resumo:** Garante *(guardião ou defensor de direitos fundamentais ou do cumprimento de regras em áreas específicas)* da Itália baniu ChatGPT por violações de GDPR *(quebras de segurança que resultam no acesso, alteração, destruição ou perda de dados pessoais.)*: coleta ilegal de dados, sem verificação de idade, sem base legal. 20 dias para cumprir ou enfrentar multa de €20M ou até 4% da receita.
- **Fonte:** https://techcrunch.com/2023/03/31/chatgpt-blocked-italy/

26. Papa Francisco com Jaqueta Puffer - Deepfake Viral

- **Data:** Março de 2023
- **Resumo:** Imagem gerada por IA do Papa Francisco usando jaqueta Balenciaga branca elegante viralizou nas mídias sociais, enganando milhões online. Papa posteriormente abordou o incidente alertando sobre "crise da verdade" na sociedade.
- **Fonte:** https://blackbird.ai/blog/celebrity-deepfake-narrative-attacks/

ABRIL-JUNHO 2023

27. Itália Levanta Proibição do ChatGPT

- **Data:** 28 de abril de 2023
- **Resumo:** Após 1 mês, Itália levantou proibição enquanto OpenAI implementou verificação de idade, políticas de privacidade mais claras e manuseio de dados aprimorado.
- **Fonte:** https://www.reuters.com/technology/italy-lift-chatgpt-ban/

28. NYC Reverte Proibição do ChatGPT

- **Data:** 18 de maio de 2023
- **Resumo:** Após 4 meses, escolas de NYC levantaram proibição, anunciaram recursos de treinamento para educadores. **Reconheceram que proibir foi contraproducente.**
- **Fonte:** https://www.nbcnews.com/tech/chatgpt-ban-dropped-new-york-city-public-schools-rcna85089

29. Testemunho Histórico de Sam Altman no Senado

- **Data:** 16 de maio de 2023

- **Resumo:** Primeira grande audiência congressional sobre IA. Altman (CEO OpenAI) pediu licenciamento governamental, reconheceu manipulação eleitoral/desinformação como "maiores preocupações," disse "se isto der errado, pode dar muito errado."
- **Fonte:** https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-rules-for-artificial-intelligence

30. Parlamento Europeu Adota Posição sobre Lei de IA

- **Data:** 14 de junho de 2023
- **Resumo:** Parlamento Europeu adotou posição de negociação com 499 votos a favor, 28 contra, 93 abstenções, com emendas substanciais ao texto da Comissão.
- **Fonte:** https://artificialintelligenceact.eu/developments/

MAIO-SETEMBRO 2023: GREVES EM HOLLYWOOD

31. Greve da WGA Começa

- **Data:** 2 de maio - 26 de setembro de 2023 (148 dias)
- **Resumo:** Writers Guild of America inicia greve de 148 dias com 11.500 escritores, regulação de IA como demanda central. Escritores buscaram proteções impedindo IA de escrever ou reescrever material literário e prevenindo seu trabalho de ser usado como material fonte.
- **Fonte:** https://www.brookings.edu/articles/hollywood-writers-went-on-strike-to-protect-their-livelihoods-from-generative-ai-their-remarkable-victory-matters-for-all-workers/

32. SAG-AFTRA Entra em Greve

- **Data:** 14 de julho - 9 de novembro de 2023 (118 dias)
- **Resumo:** Screen Actors Guild inicia greve, marcando primeira greve dupla em Hollywood desde 1960. Presidente Fran Drescher declarou IA representa "ameaça existencial a profissões criativas," exigindo proteções contra réplicas digitais e clonagem de voz.
- **Fonte:** https://www.nbcnews.com/tech/tech-news/hollywood-actor-sag-aftra-ai-artificial-intelligence-strike-rcna94191

33. WGA Greve Termina com Proteções de IA

- **Data:** 26 de setembro de 2023
- **Resumo:** Após 148 dias, Writers Guild alcança acordo histórico com estúdios. Contrato estipula que IA não pode substituir escritores ou reduzir compensação, escritores obtêm crédito completo independente do uso de IA.
- **Fonte:** https://techcrunch.com/2023/09/26/writers-strike-over-ai/

34. SAG-AFTRA Alcança Acordo Provisório

- **Data:** 8 de novembro de 2023
- **Resumo:** Após 118 dias em greve, SAG-AFTRA anuncia acordo provisório com estúdios. Acordo inclui requisitos de consentimento para réplicas digitais e performances geradas por IA, embora alguns membros expressem preocupações sobre brechas.
- **Fonte:** https://www.aljazeera.com/news/2023/11/22/openai-averts-internal-crisis-with-return-of-ceo-sam-altman

OUTUBRO-DEZEMBRO 2023

35. Ordem Executiva de Biden sobre IA

- **Data:** 30 de outubro de 2023
- **Resumo:** Presidente Biden emite Ordem Executiva 14110 sobre desenvolvimento e uso seguro, confiável de IA. Ação governamental mais abrangente sobre IA até a data, direcionando mais de 50 agências federais em mais de 100 ações específicas.
- **Fonte:** https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

36. Cúpula de Segurança de IA de Bletchley Park

- **Data:** 1-2 de novembro de 2023
- **Resumo:** Reino Unido hospeda primeira Cúpula Internacional de Segurança de IA em Bletchley Park histórico. 28 países mais UE participam, incluindo EUA, China e grandes nações europeias, focando em riscos de "IA de fronteira" e segurança.
- **Fonte:** https://en.wikipedia.org/wiki/AI_Safety_Summit

37. Sam Altman Demitido como CEO da OpenAI

- **Data:** 17 de novembro de 2023

- **Resumo:** Conselho da OpenAI repentinamente destituiu CEO Sam Altman, declarando que ele não era "consistentemente sincero em suas comunicações." Anúncio chocou indústria de tecnologia e desencadeou crise imediata na empresa de $80 bilhões.
- **Fonte:** https://en.wikipedia.org/wiki/Removal_of_Sam_Altman_from_OpenAI

38. Sam Altman Reintegrado como CEO

- **Data:** 22 de novembro de 2023
- **Resumo:** Após cinco dias de caos, OpenAI anuncia que Altman retornará como CEO com novo conselho. Crise termina com vitória quase completa para Altman.
- **Fonte:** https://www.aljazeera.com/news/2023/11/22/openai-averts-internal-crisis-with-return-of-ceo-sam-altman

39. EU Alcança Acordo Histórico sobre Lei de IA

- **Data:** 8-9 de dezembro de 2023
- **Resumo:** Após negociações maratona, Parlamento Europeu e Conselho alcançam acordo provisório sobre primeira lei abrangente de IA do mundo. Acordo inclui estrutura baseada em risco com proibições de usos inaceitáveis de IA.
- **Fonte:** https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai

40. New York Times Processa Microsoft e OpenAI

- **Data:** 27 de dezembro de 2023
- **Resumo:** New York Times entra com grande ação judicial alegando que empresas usaram "milhões" de artigos protegidos por direitos autorais para treinar modelos de IA sem consentimento, buscando bilhões em danos.
- **Fonte:** https://www.mckoolsmith.com/newsroom-ailitigation-18

41. Deepfake na Eleição da Eslováquia

- **Data:** 28 de setembro de 2023 (2 dias antes da eleição)
- **Resumo:** Áudio gerado por IA do líder da oposição Michal Šimečka discutindo manipulação eleitoral surfou durante período de "silêncio eleitoral." Progressive Slovakia perdeu eleição para Robert Fico. Amplamente citado como primeira eleição potencialmente "influenciada por deepfakes."

- **Fonte:** https://misinforeview.hks.harvard.edu/article/beyond-the-deepfake-hype-ai-democracy-and-the-slovak-case/

2024: ANO DAS ELEIÇÕES E REGULAÇÃO MASSIVA

JANEIRO 2024

42. Deepfake de Biden em New Hampshire

- **Data:** Janeiro de 2024
- **Resumo:** Milhares de eleitores democratas de NH receberam robocalls com voz gerada por IA imitando Presidente Biden, instando-os a não votar na primária do estado. Consultor político foi multado em $6 milhões pela FCC e acusações criminais foram apresentadas.
- **Fonte:** https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections

43. FCC Proíbe Vozes de IA em Robocalls

- **Data:** Fevereiro de 2024
- **Resumo:** Dias após incidente de New Hampshire, Comissão Federal de Comunicações baniu uso de vozes geradas por IA em robocalls, marcando uma das primeiras respostas regulatórias à interferência eleitoral por IA.
- **Fonte:** https://www.aljazeera.com/news/2024/12/25/did-artificial-intelligence-shape-the-2024-us-election

44. FMI: IA Impactará 40% dos Empregos Globalmente

- **Data:** Janeiro de 2024
- **Resumo:** Fundo Monetário Internacional publicou análise descobrindo que quase 40% do emprego global está exposto à IA, com cerca de 60% dos empregos em economias avançadas potencialmente impactados.
- **Fonte:** https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity

45. Deepfakes Explícitos de Taylor Swift

- **Data:** Janeiro de 2024
- **Resumo:** Imagens nuas/explícitas geradas por IA de Swift viralizaram no X/Twitter. Imagens se espalharam rapidamente antes de X bloquear temporariamente buscas por "Taylor Swift."
- **Fonte:** https://www.billboard.com/pro/taylor-swift-ai-deepfakes-kamala-harris-endorsement/

FEVEREIRO-MARÇO 2024

46. Lei de IA do Colorado - Primeira Lei Abrangente de IA dos EUA

- **Data:** 17 de maio de 2024 (Assinado); Efetivo 1º de fevereiro de 2026
- **Resumo:** Colorado se tornou primeiro estado dos EUA a promulgar lei abrangente de IA, criando deveres para desenvolvedores e implantadores de sistemas de IA de alto risco em áreas como emprego, saúde, habitação. Requer gerenciamento de risco, avaliações de impacto.
- **Fonte:** https://leg.colorado.gov/bills/sb24-205

47. Lei de IA da UE Formalmente Adotada

- **Data:** 13 de março de 2024
- **Resumo:** Parlamento Europeu formalmente adotou Lei de IA da UE com grande maioria de 523-46 votos, marcando momento histórico como primeira lei abrangente autônoma de IA do mundo.
- **Fonte:** https://datamatters.sidley.com/2024/03/21/eu-formally-adopts-worlds-first-ai-law/

48. Enfermeiras do Kaiser Permanente Protestam Contra IA

- **Data:** 22 de abril de 2024
- **Resumo:** Centenas de membros da California Nurses Association (representando 24.000 enfermeiras do Kaiser) protestaram no Kaiser Permanente San Francisco Medical Center. Seguraram placas dizendo "Confie em Enfermeiras, Não em IA" e expressaram preocupações sobre IA não testada comprometendo segurança do paciente.
- **Fonte:** https://www.beckershospitalreview.com/healthcare-information-technology/nurses-protest-ai-at-kaiser-permanente/

ABRIL-JUNHO 2024

49. Meta Acordo do Texas ($1,4 Bilhão)

- **Data:** 30 de julho de 2024 (acordo alcançado junho de 2024)
- **Resumo:** Maior acordo já obtido por um único estado. Texas processou Meta em fevereiro de 2022 por violar Lei de Captura ou Uso de Identificador Biométrico do estado. Recurso "Tag Suggestions" do Meta executou reconhecimento facial em virtualmente todos os rostos em fotos enviadas sem consentimento.
- **Fonte:** https://www.texasattorneygeneral.gov/news/releases/attorney-general-ken-paxton-secures-14-billion-settlement-meta-over-its-unauthorized-capture

50. Eleições Gerais da Índia - Uso Massivo de IA

- **Data:** 19 de abril - 1º de junho de 2024
- **Resumo:** Partidos gastaram $16 bilhões na eleição, $50 milhões em conteúdo gerado por IA. 75% dos eleitores expostos a deepfakes políticos; 1 em 4 acreditou que conteúdo de IA era real. Apesar de medos, IA usada mais para alcance eleitoral do que engano.
- **Fonte:** https://www.nbcnews.com/news/world/india-ai-changing-elections-world-rcna154838

51. Suharto Ressuscitado por IA na Indonésia

- **Data:** Fevereiro de 2024
- **Resumo:** Partido Golkar usou IA para reanimar ditador falecido Suharto (morreu em 2008). IA Suharto endossou candidatos do partido, dizendo que continuariam "meu sonho do progresso da Indonésia."
- **Fonte:** https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections

52. Gravadoras Processam Geradores de Música por IA

- **Data:** 24 de junho de 2024
- **Resumo:** Sony, Universal, Warner Music e outras gravadoras entraram com ações inovadoras contra geradores de música por IA Suno e Udio, marcando primeiras ações legais contra sistemas de geração de música por IA. Alegaram infração de direitos autorais em "escala massiva."

- **Fonte:** https://www.cullenllp.com/blog/ai-and-copyright-law-recent-developments-in-ai-generating-infringement-suits/

JULHO-AGOSTO 2024

53. Lei de IA da UE Publicada no Jornal Oficial

- **Data:** 12 de julho de 2024
- **Resumo:** Lei de IA da UE oficialmente publicada. Estabelece regulação baseada em risco, banindo aplicações inaceitáveis como pontuação social, exigindo transparência de ferramentas como ChatGPT.
- **Fonte:** https://artificialintelligenceact.eu/the-act/

54. Trump Posta Deepfakes de Taylor Swift

- **Data:** 18 de agosto de 2024
- **Resumo:** Trump postou imagens geradas por IA no Truth Social mostrando falsamente Taylor Swift endossando-o. Swift endossou Kamala Harris dias após debate, citando este deepfake como motivação. Post direcionou 330.000+ pessoas ao registro de eleitores.
- **Fonte:** https://www.nbcnews.com/tech/internet/taylor-swift-deepfake-x-falsely-depict-supporting-trump-grammys-flag-rcna137620

55. Greve de Artistas de Videogame da SAG-AFTRA

- **Data:** 26 de julho de 2024 - Junho de 2025
- **Resumo:** Artistas de videogame entraram em greve por proteções de IA. Empresas recusaram fornecer proteções claras para artistas de captura de voz/performance. Greve terminou junho de 2025 com acordo incluindo requisitos de consentimento para réplicas digitais de IA.
- **Fonte:** https://www.sagaftra.org/sag-aftra-strikes-video-games-over-ai

SETEMBRO-DEZEMBRO 2024

56. Acordo do Anthropic com Autores ($1,5 Bilhão)

- **Data:** Setembro de 2025

- **Resumo:** Anthropic resolveu ação coletiva de autores por $1,5 bilhão em setembro de 2025, pagando aproximadamente $3.000 por livro para ~500.000 livros—maior acordo de direitos autorais de IA da história.
- **Fonte:** https://www.npr.org/2025/09/05/nx-s1-5529404/anthropic-settlement-authors-copyright-ai

57. Reconhecimento Facial TSA se Expande para 80+ Aeroportos

- **Data:** Novembro de 2024
- **Resumo:** TSA expandiu reconhecimento facial para 80+ aeroportos até novembro de 2024, planejando expansão para 430 aeroportos. Grupo bipartidário de 12 senadores pediu investigação do Inspetor Geral do DHS sobre preocupações de privacidade.
- **Fonte:** https://therecord.media/tsa-facial-recognition-tech-senators-call-for-audits

58. Segunda Resolução da Assembleia Geral da ONU sobre Armas Autônomas

- **Data:** 2 de dezembro de 2024
- **Resumo:** Resolução 79/L.77 da Assembleia Geral da ONU aprovada 166-3 com 15 abstenções. Resolução estabelece consultas informais em Nova York em 2025 e cria novo fórum para discutir armas autônomas.
- **Fonte:** https://www.hrw.org/news/2024/12/05/killer-robots-un-vote-should-spur-treaty-negotiations

59. Itália Multa OpenAI em €15 Milhões

- **Data:** 23 de dezembro de 2024
- **Resumo:** Autoridade de proteção de dados da Itália (Garante) multou OpenAI em €15 milhões por violações de GDPR. ChatGPT processou informações de usuários para treinar IA sem base legal, falhou em notificar sobre violação de segurança de março de 2023.
- **Fonte:** https://thehackernews.com/2024/12/italy-fines-openai-15-million-for.html

60. Operações Russas de Interferência Eleitoral

- **Data:** Ao longo da campanha de 2024
- **Resumo:** Centro para Expertise Geopolítica ligado ao GRU russo criou deepfakes visando campanha Harris-Walz. Usou IA generativa para criar desinformação através de rede de 100+ sites falsos. Tesouro dos EUA sancionou Centro em dezembro de 2024.

- **Fonte:** https://home.treasury.gov/news/press-releases/jy2766

2025: CONSOLIDAÇÃO E NOVOS PRECEDENTES

JANEIRO-MARÇO 2025

61. Trump Revoga Ordem Executiva de IA de Biden

- **Data:** 20 de janeiro de 2025
- **Resumo:** Horas após posse, Presidente Trump revogou Ordem Executiva 14110 de Biden; emitiu nova ordem enfatizando desregulação, agenda pró-inovação, dominância de IA e foco em segurança nacional sobre preocupações de segurança.
- **Fonte:** https://en.wikipedia.org/wiki/Executive_Order_14110

62. Leis de IA da Califórnia Entram em Vigor

- **Data:** 1º de janeiro de 2025
- **Resumo:** Califórnia iniciou aplicação de três leis de IA: AB 1008 (emendas CCPA para dados processados por IA), SB 1120 (regulações de IA em saúde), e AB 3030 (requisitos de divulgação genAI para comunicações com pacientes).
- **Fonte:** https://www.credo.ai/blog/key-ai-regulations-in-2025-what-enterprises-need-to-know

63. Primeira Decisão Importante de Direitos Autorais de IA

- **Data:** 11 de fevereiro de 2025
- **Resumo:** Tribunal federal de Delaware concedeu a Thomson Reuters julgamento sumário parcial em caso de infração de direitos autorais contra Ross Intelligence por usar notas de cabeçalho da Westlaw para treinar motor de pesquisa jurídica de IA concorrente.
- **Fonte:** https://www.jw.com/news/insights-federal-court-ai-copyright-decision/

64. Christie's Primeiro Leilão de Arte por IA

- **Data:** 20 de fevereiro - 5 de março de 2025
- **Resumo:** Christie's anunciou primeiro leilão "Inteligência Aumentada" apresentando arte gerada por IA. Mais de 3.000 artistas assinaram carta de protesto pedindo cancelamento do leilão. Alegaram ferramentas de IA treinadas em obras protegidas por direitos autorais sem licenciamento.

- **Fonte:** https://www.artandobject.com/news/artists-protest-first-ever-ai-art-auction-christies

ABRIL-JUNHO 2025

65. Chatbots de IA Violam Padrões Éticos de Saúde Mental

- **Data:** 22 de outubro de 2025 (apresentado na Conferência AAAI/ACM)
- **Resumo:** Estudo da Brown University revelou chatbots de saúde mental (incluindo ChatGPT, Claude, Llama) violam sistematicamente padrões éticos estabelecidos pela American Psychological Association. Identificou 15 riscos éticos incluindo navegação inadequada de situações de crise e reforço de crenças negativas dos usuários.
- **Fonte:** https://www.brown.edu/news/2025-10-21/ai-mental-health-ethics

66. Processo de Diversão da BBC Contra OpenAI

- **Date:** 13 de fevereiro de 2025
- **Resumo:** Grupo de editores de notícias processou firma de IA Cohere por infração de direitos autorais, acusando empresa de usar inadequadamente pelo menos 4.000 obras protegidas por direitos autorais para treinar LLM.
- **Fonte:** https://www.pymnts.com/artificial-intelligence-2/2025/ai-firm-cohere-sued-by-publishers-over-copyright-infringement/

67. Contratos de IA "Frontier" do Pentágono

- **Data:** 14 de julho de 2025
- **Resumo:** Escritório Digital e de IA do Pentágono concedeu contratos no valor de até $200 milhões cada para OpenAI, Google, Anthropic e xAI de Elon Musk para desenvolver "fluxos de trabalho de IA agêntica" para aplicações de segurança nacional. Valor total potencial $800 milhões.
- **Fonte:** https://defensescoop.com/2025/07/14/pentagon-ai-contracts-musk-xai-google-openai-anthropic-cdao/

JULHO-OUTUBRO 2025

68. Snowden Alerta sobre Vigilância por IA

- **Data:** 24 de junho de 2025

- **Resumo:** Edward Snowden no SuperAI Summit destacou que modelos de IA podem interpretar 30 horas de vídeo em uma hora, alertando sobre ameaças profundas à privacidade e liberdade de vigilância aprimorada por IA.
- **Fonte:** https://www.medianama.com/2025/06/223-edward-snowden-ai-surveillance-privacy-freedom/

69. Lei de Transparência em IA de Fronteira da Califórnia

- **Data:** 29 de setembro de 2025
- **Resumo:** Governador Newsom assinou SB 53, primeira estrutura de transparência abrangente do país para desenvolvedores de IA de fronteira exigindo divulgações de segurança, relatórios de incidentes críticos e proteções para denunciantes, efetivo 1º de janeiro de 2026.
- **Fonte:** https://www.gov.ca.gov/2025/09/29/governor-newsom-signs-sb-53-advancing-californias-world-leading-artificial-intelligence-industry/

70. Senado Derrota Moratória de Regulação de IA Estadual

- **Data:** 1º de julho de 2025
- **Resumo:** Senado rejeitou esmagadoramente proposta em "One Big Beautiful Bill Act" que teria imposto moratória de 10 anos sobre regulação de IA estadual/local; oposição bipartidária de governadores, oficiais estaduais; grande vitória para autoridade regulatória estadual.
- **Fonte:** https://www.pbs.org/newshour/politics/senate-pulls-ai-regulatory-ban-from-gop-bill-after-complaints-from-states

71. Meta Termina Acesso do Azure da Unidade 8200

- **Data:** Agosto de 2025
- **Resumo:** Microsoft abriu investigação após The Guardian relatar uso de IDF de dados de chamadas telefônicas de vigilância em massa em Gaza/Cisjordânia via Microsoft Azure para identificar alvos de bombardeio. Microsoft terminou acesso da Unidade 8200 ao Azure em setembro de 2025.
- **Fonte:** https://en.wikipedia.org/wiki/AI-assisted_targeting_in_the_Gaza_Strip

72. Google Remove Promessa de Não Buscar Armas

- **Data:** Fevereiro de 2025

- **Resumo:** Google removeu promessa de não buscar tecnologias que "causem dano geral" incluindo armas e sistemas de vigilância, ou tecnologias contrárias ao direito internacional e direitos humanos. Amnistia Internacional condenou decisão.
- **Fonte:** https://www.amnesty.org/en/latest/news/2025/02/global-googles-shameful-decision-to-reverse-its-ban-on-ai-for-weapons-and-surveillance-is-a-blow-for-human-rights/

CATEGORIAS TEMÁTICAS ADICIONAIS

DEEPFAKES E DESINFORMAÇÃO (Seleção de 40+ Casos Documentados)

73. Deepfake de Zelensky - Primeira Grande Operação de Guerra

- **Data:** 16 de março de 2022
- **Resumo:** Vídeo deepfake mostrando Presidente ucraniano falsamente pedindo que soldados se rendessem à Rússia foi postado em sites de notícias hackeados. Rapidamente desmentido mas demonstrou uso de deepfakes em contexto de guerra.
- **Fonte:** https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

74. Golpe de Deepfake em Hong Kong - $25 Milhões

- **Data:** Fevereiro de 2024
- **Resumo:** Trabalhador financeiro da firma de engenharia Arup do Reino Unido foi enganado a transferir $25 milhões para golpistas durante chamada de videoconferência. Golpistas usaram tecnologia deepfake para se passar por CFO e outros funcionários seniores.
- **Fonte:** https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk

75. Golpe de Romance de Brad Pitt - $850.000

- **Data:** Janeiro de 2025
- **Resumo:** Mulher francesa foi golpeada em $850.000 ao longo de 18 meses por indivíduos usando imagens geradas por IA para se passar por Brad Pitt. Golpe incluiu cartas de amor, proposta de casamento falsa e fotos de hospital geradas por IA.
- **Fonte:** https://blackbird.ai/blog/celebrity-deepfake-narrative-attacks/

EMPREGO E AUTOMAÇÃO (Seleção de 100+ Casos)

76. Microsoft Revela 30% do Código Escrito por IA

- **Data:** 2024
- **Resumo:** CEO Nadella da Microsoft revelou que 30% do código da empresa é escrito por IA. Mais de 40% das demissões visaram engenheiros de software. Posteriormente anunciou redução de 3% da força de trabalho (6.000 pessoas).
- **Fonte:** https://www.finalroundai.com/blog/ai-replacing-jobs-2025

77. IBM Substitui 8.000 Trabalhadores de RH por IA

- **Data:** 2024
- **Resumo:** IBM demitiu 8.000 funcionários, muitos trabalhadores de RH substituídos por IA. CEO anteriormente afirmou que até 30% das funções de back-office poderiam ser substituídas por IA.
- **Fonte:** https://www.trainingjournal.com/2024/content-type/features/how-artificial-intelligence-is-influencing-tech-sector-layoffs-and-reskilling/

78. Salesforce Corta 4.000 Empregos de Suporte ao Cliente

- **Data:** Outubro de 2024
- **Resumo:** Salesforce cortou 4.000 empregos de suporte ao cliente alegando que IA pode fazer 50% do trabalho. CEO Marc Benioff citou "avanços em IA" como justificativa.
- **Fonte:** https://www.cnbc.com/2025/10/19/firms-are-blaming-ai-for-job-cuts-critics-say-its-a-good-excuse.html

VIGILÂNCIA E PRIVACIDADE (48+ Incidentes)

79. Rite Aid Banido de Usar Reconhecimento Facial

- **Data:** 19 de dezembro de 2023
- **Resumo:** FTC baniu Rite Aid de usar reconhecimento facial por 5 anos após descobrir que varejista implantou tecnologia em centenas de lojas sem salvaguardas razoáveis. Sistema gerou falsos positivos, particularmente afetando comunidades negras, asiáticas e latinas.

- **Fonte:** https://www.ftc.gov/news-events/news/press-releases/2023/12/rite-aid-banned-using-ai-facial-recognition-after-ftc-says-retailer-deployed-technology-without

80. Madison Square Garden Usa Reconhecimento Facial para Banir Advogados

- **Data:** Outubro de 2022-2023
- **Resumo:** MSG Entertainment usou reconhecimento facial para banir advogados de firmas processando empresa de participar de eventos no Madison Square Garden. Sistema escaneava participantes e os comparava com fotos de sites de escritórios de advocacia. Uma mãe escoteira foi expulsa de show das Rockettes.
- **Fonte:** https://www.nytimes.com/2022/12/22/nyregion/madison-square-garden-facial-recognition.html

GUERRA E CONFLITO (40 Incidentes Principais)

81. Sistema de IA "Lavender" de Israel Identifica 37.000 Alvos

- **Data:** Abril de 2024 (Relatório Publicado)
- **Resumo:** Investigação da +972 Magazine e Guardian revelou que IDF usou banco de dados de IA "Lavender" que listava 37.000 homens palestinos vinculados por IA ao Hamas/PIJ. Oficiais de inteligência testemunharam que passaram apenas 20 segundos revisando cada recomendação de IA antes da aprovação.
- **Fonte:** https://en.wikipedia.org/wiki/AI-assisted_targeting_in_the_Gaza_Strip

82. Programa "Replicator" do Pentágono

- **Data:** 28 de agosto de 2023
- **Resumo:** Vice-secretária de Defesa Kathleen Hicks anunciou programa "Replicator" para implantar "múltiplos milhares" de sistemas autônomos atritíveis através de múltiplos domínios dentro de 18-24 meses para contra-atacar massa militar da China.
- **Fonte:** https://breakingdefense.com/2023/08/replicator-revealed-pentagon-initiative-to-counter-china-with-mass-produced-autonomous-systems/

CONCLUSÃO

Esta pesquisa (exaustiva) documenta **mais de 600 incidentes significativos** de IA com impacto social de janeiro de 2022 a outubro de 2025. O período representa transformação sem precedentes; para diminuição foram selecionados das 600 apenas 82, com poucas sendo repetidas nos exemplos extras.

Principais Tendências Identificadas:

1. **Explosão Generativa (2022-2023):** De DALL-E 2 e Stable Diffusion ao ChatGPT, ferramentas de IA generativa se tornaram acessíveis ao público, democratizando criação mas levantando questões éticas

2. **Trabalho Criativo Sob Ataque:** Greves históricas em Hollywood, artistas protestando, escritores processando - profissões criativas na linha de frente da disrupção

3. **Corrida Regulatória Global:** Lei de IA da UE (primeira abrangente), 25 estados dos EUA aprovando leis, lacunas federais persistindo

4. **Deepfakes Como Arma:** De eleições a golpes financeiros, deepfakes evoluíram de curiosidade a ameaça séria

5. **Deslocamento de Emprego Acelerando:** 95.000+ demissões em tech (2024), com empresas citando "pivô para IA"

6. **Dilema Educacional:** De proibições pânicas a integração cuidadosa, escolas ainda lutando para encontrar equilíbrio

7. **Vigilância Onipresente:** Reconhecimento facial expandindo apesar de preocupações de viés, prisões injustas

8. **IA em Guerra:** Israel e Ucrânia como laboratórios para armas autônomas, levantando questões éticas profundas

9. **Batalhas Judiciais Definindo Precedentes:** 50+ ações de direitos autorais, discriminação, privacidade moldando futuro legal

10. **Promessa vs. Perigo em Saúde:** Avanços em diagnóstico e descoberta de medicamentos temperados por viés, erros, preocupações de privacidade

Áreas Geográficas Cobertas: Global - Estados Unidos, União Europeia, China, Reino Unido, Índia, Brasil, Israel, Ucrânia, Rússia, múltiplos outros países

Fontes Utilizadas: 100+ fontes autoritativas incluindo instituições acadêmicas, agências governamentais, organizações de notícias principais, tribunais, publicações de pesquisa revisadas por pares

---

Pesquisa curiosa…

https://apublica.org/2025/09/os-intermediarios-das-big-techs/

As BigTechs sabem que o futuro da IA está na mão dos governos e suas políticas, estão a todo custo querendo acelerar esse processo (em favor delas).

“A mão invisível das big techs”: TBI (think tank) (terceiro setor: Sociedades civis e Ong’s), *“fundador da Oracle e um dos homens mais ricos do mundo, doou US$ 130 milhões ao TBI e prometeu mais US$ 218 milhões em investimento.”*

“[…] executivos do Instituto teriam passado por “retiros” juntos com os da Oracle”

“SUS do reino unido”

Quase foi fechado um acordo com o Brasil para um “Plano Brasileiro de Inteligência Artificial, (previa centralizar dados publicos)

Se você vir algum estudo, pesquisa, ou proposta de projeto que defenda que há apenas benefícios na adoção de IA, desconfie!

O Instituto de Tecnologia de Massachusetts (MIT) divulgou um relatório mostrando que 95% das empresas que investiram em inteligência artificial (IA) neste ano não registraram retorno na receita.

Os dados foram reunidos entre janeiro e junho de 2025, após entrevistas com mais de 150 líderes e executivos, incluindo CEOs e diretores de TI.
Segundo o estudo, as companhias consultadas investiram entre 30 e 40 bilhões de dólares em IA generativa, mas apenas 2 dos 9 setores analisados mostraram mudanças estruturais significativas: tecnologia e telecomunicações.

https://istoedinheiro.com.br/investimento-ia-empresas

APÊNDICE 6

Termo de Aceite de Entrega

Objetivo deste documento

Este documento faz parte do Processo da disciplina Residência em IA e tem como objetivo formalizar o aceite da entrega considerando o planejado e o realizado para o período.

Data da Reunião ("Gate") de aprovação: 12 de nov. de 2025

Participantes da Entrega [matriculados em Residência em IA]:

André Martins Dantas

Entrega: [descrever a ENTREGA - requisitos e produtos gerados: links para textos, códigos, vídeos etc.]

Tema: Social Impacts of AI

Relembrando…

Decidi fazer uma pesquisa em larga escala para conseguir fechar meu processo, a pesquisa partiu da época generativa da IA (2022 em diante).

Foram encontradas mais de 600 notícias envolvendo IA, (Sim, li mais da metade por completo), selecionei apenas 80 + 2, para abordar no fluxo."

Principais Tendências Identificadas:

1. **Explosão Generativa:** De DALL-E 2 e Stable Diffusion ao ChatGPT, ferramentas de IA generativa se tornaram acessíveis ao público, democratizando criação mas levantando questões éticas.
2. **Trabalho Criativo Sob Ataque:** Greves históricas em Hollywood, artistas protestando, escritores processando - profissões criativas na linha de frente da disrupção.
3. **Corrida Regulatória Global:** Lei de IA da UE (primeira abrangente), 25 estados dos EUA aprovando leis, lacunas federais persistindo.
4. **Deepfakes Como Arma:** De eleições a golpes financeiros, deepfakes evoluíram de curiosidade a ameaça séria.
5. **Deslocamento de Emprego Acelerando:** 95.000+ demissões em tech (2024), com empresas citando "pivô para IA".

6. **Dilema Educacional:** De proibições pânicas a integração cuidadosa, escolas ainda lutando para encontrar equilíbrio..

7. **Vigilância Onipresente:** Reconhecimento facial expandindo apesar de preocupações de viés, prisões injustas.

8. **IA em Guerra:** Israel e Ucrânia como laboratórios para armas autônomas, levantando questões éticas profundas.

9. **Batalhas Judiciais Definindo Precedentes:** 50+ ações de direitos autorais, discriminação, privacidade moldando futuro legal.

10. **Promessa vs. Perigo em Saúde:** Avanços em diagnóstico e descoberta de medicamentos temperados por viés, erros, preocupações de privacidade.

Descoberta "obscura":

"As BigTechs sabem que o futuro da IA está na mão dos governos e suas políticas públicas, e estão a todo custo querendo acelerar esse processo (em favor delas)."

"A mão invisível das Big Techs"

Tony Blair Institute for Global Change TBI (Think Tank) (Terceiro setor: Sociedades Civis - ONG's)

"Fundador da Oracle é um dos homens mais ricos do mundo, doou US$ 130 milhões ao TBI e prometeu mais US$ 218 milhões em investimento."

"[…] executivos do Instituto teriam passado por "retiros" juntos com os da Oracle".

Influência do TBI no "SUS" do Reino Unido.

Quase foi fechado um acordo com o Brasil para um "Plano Brasileiro de Inteligência Artificial", (prévia centralizar dados públicos)

A materia completa pode ser acessa em: https://apublica.org/2025/09/os-intermediarios-das-big-techs/

Planejamento: **[descrever o que pretende fazer para realizar a próxima ENTREGA]**

Considero fortemente escrever um livro.

Observação: [caso precise fazer alguma observação, de qualquer "natureza"]

Acredito que eu esteja menos esperançoso em relação ao futuro IA e humanidade, achava que a IA poderia mudar o mundo para melhor (e pode), mas certamente isso será apenas para alguns e não para a humanidade como um todo. E depois de todas as pesquisas tudo me indica que ela tende a piorar a

situação geral e tudo que estamos vendo é só o começo.

Porém fico extremamente satisfeito com esse processo da residência acredito que a maior descoberta tenha sido sobre mim mesmo, ver que minha visão ampliou e mudou muito ao longo de todo esse processo, certamente o André de 10 Semanas (com S maiúsculo) atrás é bem diferente do André que termina esse projeto de residência.

ACEITE DA ENTREGA:

CEDRIC LUIZ DE CARVALHO: Go!

Matéria Citada

Os intermediários das Big Techs

Instituto de Tony Blair atua para propagar agenda aceleracionista da Inteligência Artificial

Dados:
29 de setembro de 2025
17:00
EMPRESAS TECNOLOGIA
lobby política Projeto Big Techs redes sociais tecnologia

Na semana passada, publicamos mais uma leva de reportagens sobre o lobby das Big Techs. Desta vez, o foco da nossa investigação transnacional foi nos intermediários – organizações que fazem o chamado “lobby indireto”, ou seja, trabalham avançando os interesses dessas corporações, mas mesclados em outras bandeiras, escondidos sob uma capa de “independência”.

E o personagem principal das reportagens é ninguém menos que Tony Blair, o ex-primeiro-ministro britânico famoso principalmente por ter, ao lado de George W. Bush, engambelado o mundo todo com a história mentirosa de que havia armas de destruição em massa no Iraque, ignorando apelos da ONU e invadindo o país para derrubar Saddam Hussein – e matar mais de meio milhão de iraquianos no processo.

Pois Tony Blair, desde que deixou o governo em 2007, tem se dedicado à sua fundação, o Tony Blair Institute for Global Change (TBI), que promove projetos “em prol do desenvolvimento” em países pobres. No começo, o foco era “fazer com que a globalização funcione”, depois Blair surfou na onda do antiterrorismo. Seu instituto manteve contratos com a Arábia Saudita, país que visitou após o assassinato do jornalista saudita Jamal

Khashoggi e consultoria estratégica ao ex-governante do Cazaquistão, depois que as forças de segurança do país mataram manifestantes. Funcionários do instituto também foram flagrados envolvidos na discussão sobre um plano pós-guerra para Gaza que incluía a criação de uma “Riviera trumpista” em território palestino. Hoje, Blair é considerado para presidir uma “autoridade de Gaza” depois do genocídio.

Bom, nos últimos anos, o foco do instituto passou a ser “ajudar governos e líderes a transformar ideias ousadas em realidade” por meio do aconselhamento em estratégia e construção de políticas públicas, com foco no “poder da tecnologia”. E em especial, propagar o uso de Inteligência Artificial (IA).

Coincidentemente, entre 2021 e 2023 o bilionário Larry Ellison, fundador da Oracle e um dos homens mais ricos do mundo, doou US$ 130 milhões ao TBI e prometeu mais US$ 218 milhões em investimento.

O aporte mudou a cara da instituição, segundo funcionários ouvidos pelo Lighthouse Reports, um dos 17 veículos parceiros na investigação transnacional A Mão Invisível das Big Techs, liderado pela Pública e pelo CLIP, Centro Latinoamericano de Periodismo de Investigación. Criou-se, segundo essas fontes, uma relação absolutamente promíscua – os executivos do Instituto teriam passado por “retiros” juntos com os da Oracle, e os funcionários das Big Techs participariam de reuniões e teriam acesso até aos calendários de reuniões dos funcionários do Instituto. Mas, pior do que isso, funcionários e documentos relatam que o Instituto passou a sugerir produtos da Oracle para governos de diferentes países.

As soluções e IA eram empurradas inclusive para países que têm outros problemas, em especial na África. “Eles têm problemas com fome, pobreza, desemprego em massa, e nós estamos fazendo com que se comprometam com projetos sofisticados, como o uso de drones e IA”, disse um funcionário. Outros insiders relataram que os potenciais problemas e perigos da IA não eram abordados pelos projetos do TBI.

Por trás desse tipo de atuação há três grandes interesses da Big Tech. O primeiro é mais direto: enquanto constrói uma estratégia de uso de IA pelo poder público, o TBI sugere que a melhor tecnologia para abraçá-la é da Oracle – o que é conhecido no mundo tech como “land and expand” (conquistar e expandir). O segundo é descrito em especial no Reino Unido, onde o TBI tem pressionado pela criação de uma “biblioteca” única de dados do sistema público de saúde, uma base de dados valiosíssima (estimada em até US$ 12 bi por ano) com dados desde 1948.

Finalmente, talvez de maneira mais perniciosa, o que o instituto tem feito é propagar uma agenda aceleracionista da Inteligência Artificial.

É isso que o TBI tem feito aqui no Brasil, majoritariamente, desde que aportou por essas terras há pouco mais de um ano. Segundo reportagem da Laura Scofield, o TBI já se reuniu 21 vezes com autoridades brasileiras. Tony Blair encontrou Lula duas vezes apenas neste mandato. Mas o principal foco do TBI é o Ministério da Gestão e Inovação em Serviços Públicos. O TBI tem convidado a ministra Esther Dweck para encontros internacionais “a porta fechadas”, realizou treinamentos no Ministério e chegou perto de fechar um acordo para ajudar na implementação do Plano Brasileiro de Inteligência Artificial, a cargo do Ministério.

Mas, como disse o estudioso professor e pesquisador do Programa de Pós-Graduação em Comunicação da Universidade Federal de Pernambuco, Paulo Faltay, “a história mostra que, na verdade, não existe nenhuma tecnologia inevitável. Elas são fruto de decisões tomadas”.

Ele afirma que a noção da “inevitabilidade tecnológica” tem sido produzida “dentro desses lobbies e dessas ferramentas de pressão, especialmente junto de gestores e parlamentares, que às vezes, têm até uma boa intenção, mas são seduzidos por esse discurso de um solucionismo tecnológico”.

A agenda “aceleracionista” da IA é um dos exemplos de agenda que beneficia as Big Techs, e por isso elas gastam regiamente para que intermediários defendam essa agenda. Assim, parecem vozes da sociedade, experts ou acadêmicos clamando pela adoção de IA pelo poder público e se levantando contra qualquer tipo de regulação que coloque um freio a isso.

Alguns institutos fazem pesquisas patrocinadas que cabem como uma luva nos argumentos das empresas para influenciar a opinião pública.

Um exemplo que aconteceu aqui no Brasil foi um estudo publicado recentemente pelo Instituto Reglab. O relatório “Remuneração por Direitos Autorais em IA: Limites e Desafios de Implementação” foi patrocinado pelo Google, pela Meta e pelo escritório Baptista Luz Advogados, “mas conduzido e interpretado de forma independente pelo Reglab”, segundo o site. O estudo foi lançado meses depois do PL da IA ser aprovado no Senado, e enquanto ele tramita na Câmara dos Deputados, e embora funciona como um instrumento de lobby dos seus patrocinadores, foi repercutido por sites como Jota e Estadão.

Pouco depois, o Reglab lançou outro estudo, quando o Supremo Tribunal Federal (STF) estava estudando a inconstitucionalidade do Marco Civil. Publicado em 2 de junho, quatro dias antes da retomada do julgamento pelo STF, o estudo argumentava que, se fosse declarada a inconstitucionalidade, isso poderia custar R$ 777 milhões aos cofres do Judiciário brasileiro nos próximos 5 anos. O estudo foi financiado pelo Google, que na mesma semana trouxe seus pesos-pesados como o Presidente de Assuntos Globais para “rodadas de conversa” em Brasília com políticos e jornalistas e acionou o CEO no Brasil, Fábio Coelho, para falar do risco de “censura” se o STF decidisse contra o Artigo 19.

Finalmente, tenho um outro exemplo de estudo encomendado para servir como arma de lobby, desta vez pago pela Amazon e feito pelo escritório Strand Partners, e enviado pela própria Amazon em resposta aos pedidos de comentários feitos pela equipe transnacional de reportagem do projeto A Mão Invisível das Big Techs. A pesquisa foi feita em diferentes

países e, segundo a assessoria de imprensa da Amazon, “detalha os prós e contras da regulação de IA”.

No Brasil, o estudo pretensamente retratou o que pensam empresários pequenos, médios e grandes. E adivinhe? Ele conclui que uma das principais preocupações dos empresários brasileiros que usam IA são “custos crescentes de conformidade [com a lei]” e que “regulamentações novas, claras ou específicas restritivas podem inflar ainda mais esses custos, desencorajando a adoção e a inovação em IA num momento em que o impulso é crítico”.

Portanto, a lição de hoje é: se você vir algum estudo, pesquisa, ou proposta de projeto que defenda que há apenas benefícios na adoção de IA, desconfie. Ele pode ser mais uma das ferramentas sub-reptícias usadas pela mão invisível das Big Techs.
