

## **TECHLAB**

### **RELATÓRIO DO GRUPO DE APRESENTAÇÃO 4 - ESTADOS UNIDOS (28/10)**

Bruno Hernandes Leão (13637936)

Elias Esteban Sujin Kim (12509761)

Maria Luiza Cavalcanti (13738051)

Thiago Henrique Lemos Costa (12691474)

## **THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN: 2019 UPDATE**

### **ESTRATÉGIA 1 - PESQUISA**

#### **Avançar quanto às metodologias utilizadas nos dados para a descoberta de conhecimentos**

- A veracidade dos dados é particularmente difícil quando se trata de grandes quantidades de dados, o que dificulta a avaliação e extração de conhecimentos por parte dos seres humanos.
- Embora muita investigação tenha abordado veracidade através de métodos de garantia da qualidade dos dados para efetuar a limpeza dos dados e a descoberta de conhecimentos, são necessários mais estudos para melhorar a eficiência das técnicas de limpeza de dados,
- aprendizagem automática multimodal grande variedade de diferentes tipos de dados (por exemplo, discretos, contínuos, de texto, espaciais, temporais, espaço-temporais, gráficos)
- Devem também determinar a forma de identificar e processar eventos raros para além das abordagens puramente estatísticas; trabalhar com outras fontes de conhecimento (ou seja, qualquer tipo de informação que explique o mundo, como o conhecimento da lei da gravidade ou das normas sociais e obter aprendizado da IA mesmo quando há escassez de dados)

#### **Compreender as capacidades e limitações teóricas da IA**

- compreender melhor porque é que as técnicas de IA - especialmente a aprendizagem automática - funcionam frequentemente bem na prática
- Embora diferentes disciplinas (incluindo a matemática, as ciências do controlo e a informática) estão a estudar esta questão, o campo carece atualmente de modelos ou quadros teóricos unificados para compreender o desempenho dos sistemas de IA.

#### **Prosseguir as pesquisas em IA de maneira geral**

- IA geral e restrita Mas estes sistemas exemplificam uma IA limitada, uma vez que só podem ser aplicados às tarefas para as quais foram especificamente concebidos.
- Chatp GPT e outros programas utilizados atualmente mostram que essa IA já está mais flexível e autónoma

### **Desenvolver sistemas escaláveis de IA**

- Muitos esforços anteriores centraram-se em técnicas de planeamento e coordenação centralizadas. no entanto, estas abordagens estão sujeitas a pontos únicos de falha, como a perda do planejador, ou a perda da ligação de comunicações com o planejador.
- As técnicas de planeamento e controle distribuídas são mais difíceis de realizar algoritmicamente e são frequentemente menos eficientes e incompletas, mas oferecem potencialmente mas oferecem potencialmente maior robustez a pontos de falha únicos.
- A investigação futura deve descobrir técnicas mais eficientes, robustas e técnicas mais eficientes, robustas e escaláveis para o planeamento, controlo e colaboração de equipas de múltiplos sistemas de IA e seres humanos.

### **Fomentando a pesquisa na “human-like IA”**

- human-like ia é aquela que absorve os dados e tenta aprender a reproduzir ou criar através deles de uma forma similar a nossa
- No entanto, existe um fosso significativo entre a forma como os actuais algoritmos de IA funcionam e a forma como as pessoas aprendem e executam tarefas.
- As pessoas são capazes de aprender com apenas alguns exemplos, ou recebendo instruções formais e/ou "dicas" para a execução de tarefas, ou observando outras pessoas a executar essas tarefas.

### **Desenvolver robôs com maior capacidade e confiabilidade**

- robots generalizar as suas capacidades, efetuar uma autoavaliação do seu desempenho atual e aprender um repertório de movimentos físicos com professores humanos.
- quase como se ele tivesse uma personalidade mesmo
- A mobilidade e a manipulação são áreas a investigar para que os robots possam deslocar-se em terrenos acidentados e incertos e manipular uma variedade de objectos com destreza.
- Os robôs precisam de aprender a trabalhar em equipe de uma forma e colaborar com os humanos de uma forma fiável e previsível.

### **Melhorando o hardware para melhorar a eficiência da IA**

- O desenvolvimento de hardware otimizado para algoritmos de IA permitirá níveis de desempenho ainda mais elevados do que os das GPU.
- Um exemplo são os processadores "neuromórficos" que são vagamente inspirados na organização do cérebro e, em alguns casos, otimizados para o funcionamento de redes neuronais.
- Os avanços a nível do hardware podem também melhorar o desempenho dos sistemas de IA que consomem muitos dados.

## **ESTRATÉGIA 2 - DESENVOLVER MÉTODOS EFICAZES PARA A COLABORAÇÃO ENTRE HUMANO E IA**

- Embora já existam abordagens eficazes para a colaboração entre o ser humano e a IA já existem, a maior parte delas são "soluções pontuais" que só funcionam em ambientes específicos utilizando plataformas específicas para objectivos específicos.

Objetivo é criar um cenário em que

1. A IA desempenha funções ao lado do ser humano: Os sistemas de IA executam tarefas periféricas que apoiam o humano na tomada de decisões. Por exemplo, a IA pode ajudar os seres humanos com a memória de trabalho, a recuperação da memória de curto ou longo prazo e as tarefas de previsão.
2. A IA desempenha funções quando o ser humano se depara com uma elevada sobrecarga cognitiva: Os sistemas de IA desempenham funções de monitorização complexas (como os sistemas de aviso de proximidade do solo em aeronaves), a tomada de decisões e diagnósticos médicos automatizados quando o ser humano precisa de assistência.
3. A IA desempenha funções em vez de um ser humano: Os sistemas de IA executam tarefas para as quais os seres humanos têm capacidades muito limitadas, tais como operações matemáticas complexas, orientação do controlo de sistemas dinâmicos em ambientes operacionais contestados, aspectos do controlo de sistemas automatizados em ambientes nocivos ou tóxicos, e em situações em que um sistema deve responder muito rapidamente (por exemplo em salas de controlo de reactores nucleares).

Para responder a estas preocupações, devem ser utilizados certos princípios de automatização centrados no ser humano na conceção e desenvolvimento destes sistemas:56

1. Empregar uma conceção intuitiva e de fácil utilização das interfaces, controlos e ecrãs do sistema de IA humana.
2. Manter o operador informado. Apresentar informações críticas, estados do sistema de IA e alterações a esses estados
3. Manter o operador treinado. Efetuar formação recorrente sobre conhecimentos, competências e capacidades gerais (KSAs), bem como formação em algoritmos e lógica empregues pelos sistemas de IA e os modos de falha esperados do sistema.
4. Tornar a automatização flexível. A implantação de sistemas de IA deve ser considerada como uma opção de projeto para operadores que desejam decidir se querem ou não utilizá-los. Também é importante a conceção períodos de carga excessiva de trabalho

### **Em busca de novos algoritmos para uma IA com consciência humana**

- interação entre humano e máquina sem descontinuidades
- Devem ser desenvolvidos modelos de interrupção que permitam a um sistema inteligente que permitam a um sistema inteligente interromper o ser humano apenas quando necessário e adequado.

- Os sistemas inteligentes devem também ter a capacidade de aumentar a cognição humana, sabendo que informação recuperar quando o utilizador precisa dela, mesmo que não tenha solicitado explicitamente essa informação ao sistema.
- No futuro sistemas inteligentes devem ser capazes de ter em conta as normas sociais humanas e atuar em conformidade.
- Os sistemas inteligentes podem trabalhar mais eficazmente com os humanos se possuírem algum grau de inteligência emocional, de modo a para que possam reconhecer as emoções dos seus utilizadores e responder adequadamente.

### **Desenvolver técnicas de IA para aumentar a capacidade humana**

- A investigação sobre o aumento da capacidade humana inclui algoritmos que funcionam em fixos (como um computador); dispositivos portáteis (como óculos inteligentes); dispositivos implantados (como interfaces cerebrais); e em ambientes específicos do utilizador.
- assistente médico assinala um erro num procedimento médico, com base em leituras de dados combinadas de vários dispositivos

### **Desenvolver sistemas de processamento da língua mais eficazes**

- Os sistemas atuais não conseguem lidar com os desafios do mundo real, como o discurso em ambientes ruidosos, o discurso com muito sotaque, discurso de crianças, discurso de deficientes e discurso para línguas gestuais.
- Esses sistemas terão de inferir os objetivos e as intenções dos seus interlocutores humanos, utilizar o registo, o estilo e a retórica adequados à situação, e empregar estratégias de reparação em caso de mal-entendidos no diálogo.
- estado emocional, afeto e postura

### **ESTRATÉGIA 3: ENTENDER E DIRECIONAR AS IMPLICAÇÕES ÉTICAS, JURÍDICAS E SOCIAIS DA IA**

- colaboração intensiva entre peritos técnicos, bem como entre as partes interessadas e especialistas noutros
- outros domínios, incluindo as ciências sociais e comportamentais, o direito, a ética e a filosofia
- A pesquisa neste domínio pode beneficiar de perspectivas multidisciplinares que envolvam peritos de ciências informáticas, ciências sociais e comportamentais, ética, ciências biomédicas, psicologia, economia direito e investigação política.

### **Improving fairness, transparency, and accountability by design**

- Os investigadores devem aprender a conceber estes sistemas de modo a que as suas acções e decisões sejam transparentes e facilmente interpretáveis pelos seres humanos e, assim, possam ser examinadas em função de eventuais preconceitos que possam conter, em vez de se limitarem a aprender e repetir esses preconceitos.
- Há sérias questões intelectuais sobre como representar e "codificar" sistemas de valores e crenças.

### **Construindo a ética da IA**

- As questões éticas variam consoante a cultura, a religião e as crenças. No entanto, podem ser desenvolvidos quadros de referência éticos aceitáveis para orientar o raciocínio e a tomada de decisões dos sistemas de IA, a fim de explicar e justificar as suas conclusões e acções. até se basear em definições da ONU ou que a maioria dos países adotam
- Uma abordagem multidisciplinar é necessária para gerar conjuntos de dados para treino que reflectam um sistema de valores adequado, incluindo exemplos que indiquem o comportamento preferido quando confrontados com questões morais difíceis ou com valores contrad

#### **Estruturas para concretizar a ética da IA**

- Devem ser feitos progressos adicionais na investigação fundamental para determinar a melhor forma de conceber arquitecturas para sistemas de IA que incorporem o raciocínio ético.
- Foi sugerida uma variedade de abordagens, como uma arquitetura de monitor de dois níveis que separa a IA operacional de um agente monitor responsável pela avaliação ética ou legal de qualquer ação operacional.
- Uma visão alternativa é a preferência pela engenharia de segurança, na qual é utilizado um quadro conceitual preciso para a arquitetura do agente de IA para garantir que o comportamento da IA é seguro e não prejudicial para os seres humanos.
- Um terceiro método consiste em formular uma arquitetura ética utilizando princípios teóricos de conjuntos, combinados com restrições lógicas

#### **Estratégia 4: Garantir a proteção e a segurança dos sistemas de IA:**

- Aprimoramento da transparência;
- Construção de confiança;
- Aprimoramento de validação e verificação;
- Segurança contra ataques.

#### **Estratégia 5: Desenvolver conjuntos de dados e ambientes públicos partilhados para treino e teste de IA.**

- Agências devem melhorar inventário de dados;
- Objetivos:
  - Fomentar pesquisa e usabilidade
    - Priorizar melhorias no acesso de dados de qualidade sobre a AI
- FAIR (findable, accessible, interoperable, and reusable) data;
- Documentar datasets e model provenance vai melhorar a confiança e uso responsável das tecnologias IA;
- Integridade e a disponibilidade dos datasets de treinamento e teste de IA - cruciais para garantir resultados confiáveis;
- Tornar treinamentos e testes adequados aos interesses comerciais e públicos;
- Desenvolvimento de bibliotecas e kits de ferramentas de software de código aberto

### **Programas recentes**

- VA DATA COMMONS - “medical-genomical” dataset;
- Naturalistic Driving Study (NDS) - gravou 5.4 milhões de viagens de 3,400 motoristas;
- NIH Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) - acesso a dados armazenados em nuvem;

### **Estratégia 6: Medir e avaliar as tecnologias de IA através de normas e padrões de referência.**

- Parâmetros - confiabilidade dos sistemas de IA;
- Desenvolvendo um amplo espectro de padrões de IA
  - deve ser acelerado para acompanhar as capacidades em rápida evolução e expansão de domínios de aplicações de IA.
- Padrões de IA (p.37):
  - Software engineering; Performance; Métrica; Segurança; Usabilidade; Interoperabilidade; Privacidade; Rastreabilidade.
- Os benchmarks impulsionam a inovação ao promover avanços destinados a abordar cenários estrategicamente selecionados;
- Dados objetivos para acompanhar a evolução da ciência e das tecnologias de IA.
- **Envolver a comunidade de IA em padrões e benchmarks;**
- **Aumentar a disponibilidade de bancos de testes de IA.**

### **Estratégia 7: Compreender melhor as necessidades da mão de obra nacional de P&D em matéria de IA.**

- Escritório dos EUA das Estatísticas do Trabalho prevê que o número de vagas para informática e informação cientistas e engenheiros crescerão 19% de 2016 a 2026.
- **Equipes multidisciplinares** - não apenas cientistas e engenheiros da computação, mas também aplicações de inovação, ciência cognitiva e psicologia, economia e teoria dos jogos, engenharia e teoria de controle, ética, linguística, matemática, filosofia.

### **Estratégia 8: Expandir as parcerias público-privadas para acelerar os avanços na IA.**

- Crescer capacidade de pesquisa e força de trabalho, conectando professores e estudantes universitários com a indústria representantes, ambientes da indústria e empregos na indústria.
- Melhorar o desenvolvimento da força de trabalho, ampliação da participação (Multisector partnerships).
- Colaborações para implantar e melhorar a infraestrutura de pesquisa.
- Programas conjuntos para promover a investigação aberta, competitiva e fundamental.

### **Commission on Artificial Intelligence Competitiveness, Inclusion, and Innovation Report and Recommendations**

Relatório da Comissão da Competitividade, Inclusão e Inovação da Inteligência Artificial (divulgado em 2023) > U.S. Chamber of Commerce

Chamber Technology Engagement Center

Advoga por um framework regulatório baseado no risco que possibilitaria a implementação responsável e ética da IA. Prioriza a inovação. Capital Economics> EUA país com maior potencial de se beneficiar da IA

Por que regular a IA?

Regulamentar a IA é essencial para criar programas de conformidade que promovam a inovação em IA.

Isso garante transparência, explicabilidade e uso responsável dos sistemas de IA.

Prevenção de Disparidades em Saúde: A área de saúde é um exemplo primordial da necessidade de regulamentação em IA. Disparidades de saúde já são abundantes nos Estados Unidos, e a falta de regulamentação em IA pode agravá-las ao incorporar viés nos modelos de IA e na automação.

### **Abordagens para a regulação da IA**

**Desafios da Regulação:** Debate sobre a responsabilidade do governo versus entidades privadas na regulação da IA. Dilema entre fornecer transparência, explicabilidade e responsabilidade da IA e promover inovações benéficas para a sociedade.

**Equilíbrio na Regulação:** O risco de regulamentação excessiva que pode prejudicar a inovação versus regulamentação insuficiente que pode causar danos à sociedade.

**Consistência e Harmonização:** A necessidade de consistência nas regulamentações de IA para evitar conflitos e desafios de conformidade.

**Alinhamento Global na Regulação:** A importância de regulamentações de IA alinhadas globalmente para evitar custos e incertezas desnecessárias.

**Clareza e Certeza Legal:** A necessidade de clareza e certeza nas abordagens regulatórias para empresas que desenvolvem e usam IA.

**Modelo Regulatório da FDA:** Referência ao modelo regulatório da Food and Drug Administration (FDA) como exemplo para regulamentação da IA.

**Abordagem Baseada em Risco:** A tendência em direção a abordagens baseadas em risco na regulamentação da IA para equilibrar inovação e segurança.

**Envolvimento de Stakeholders:** A importância da colaboração entre empresas, defensores do consumidor, acadêmicos, sociedade civil e reguladores na elaboração de regulamentações eficazes.

**Governança Multistakeholder:** Recomendação de trabalhar com organizações e agências, como o NIST, para desenvolver um quadro de governança multistakeholder.

**Consenso Internacional na Regulação da IA:** A convergência de opiniões em países como os EUA, Reino Unido, Canadá e UE em direção a abordagens baseadas em risco para a regulamentação da IA.

**Desafios da Regulação de IA:** Questões de equidade, viés, igualdade e risco de discriminação abordados pelas regulamentações de IA.

**Desafios na Mitigação do Viés:** Foco na mitigação de danos

**Importância da Transparência:** Reconhecimento da importância da transparência na regulamentação da IA.

**Integração da Regulação de IA com a Privacidade:** Importância de que a regulamentação de IA seja informada pelas leis de proteção de dados pessoais.

**Segurança dos Dados Pessoais:** Necessidade de regras de IA que mantenham a segurança dos dados pessoais em todas as indústrias, em vez de regulamentos diferenciados.

**Necessidade de uma Lei de Privacidade Federal:** Os Estados Unidos devem promulgar uma lei de privacidade federal robusta para permitir melhores práticas de dados e desenvolvimento de sistemas de IA.

## **ADPPA - American Data Privacy and Protection Act**

**Competitividade Internacional:** Importância da regulamentação de dados para a competitividade internacional dos EUA em relação a outras regiões, como a União Europeia e China.

**Colaboração entre Legisladores, Reguladores e Empresas:** Importância de uma colaboração eficaz entre legisladores, reguladores e o setor empresarial para determinar se as regulamentações atuais são adequadas ou se novas regulamentações são necessárias.

**Ênfase na Neutralidade Tecnológica das Leis:** As leis devem ser neutras em relação à tecnologia, focando nas aplicações e resultados da IA, em vez de tecnologias específicas.

**Abordagem Baseada em Risco:** Importância de adotar uma abordagem baseada em risco para a regulamentação da IA, considerando o impacto da IA em áreas como habitação, educação, emprego e serviços de saúde.



**Com base no impacto individual:** A classificação de risco deve ser determinada com base no impacto na vida dos indivíduos e na comunidade, em vez de categorias amplas predefinidas.

**Avaliações de Impacto Obrigatórias para Usos de Alto Risco:** Necessidade de avaliações de impacto obrigatórias para determinar se um uso específico de IA deve ser classificado como de alto risco.

**Ênfase na Interoperabilidade das Regulamentações de IA em Fronteiras:** Uma abordagem baseada em risco também pode promover a interoperabilidade entre regulamentações de IA em diferentes jurisdições.

**Regras Estritas e Transparência em Casos de Alto Risco:** Necessidade de regras estritas, garantias legais e requisitos de transparência para casos de alto risco de IA.

**Evitar Proibições Categóricas, a Menos que sejam Extremas:** As proibições amplas de usos de IA não são necessárias, a menos que sejam usadas para ações ilegais ou extremas que ameacem a segurança pública.

**Diferença entre os papéis e responsabilidades dos provedores de IA e dos implementadores de IA:** Os desenvolvedores de IA criam as ferramentas, mas os implementadores de IA determinam como elas são usadas. Qualquer legislação de IA deve definir claramente as responsabilidades tanto dos provedores de IA quanto dos implementadores de IA.

**Considerar ambientes de testes regulatórios para áreas de inovação de alta prioridade:** Ambientes de testes regulatórios promovem a inovação sem prejudicá-la, especialmente para a IA considerada de alto risco, mas essencial para a competitividade dos EUA. Centros de inovação regionais podem ajudar a adaptar soluções a regiões específicas e impulsionar a inovação sustentável.

**Promover Acesso a Dados Críticos para Impulsionar a Inovação em IA:** Apoiar o Acesso Público a Conjuntos de Dados Chave para Desbloquear um Valor Maior: O uso efetivo de dados é essencial para a competição econômica dos EUA na área de IA.

**Incentivar Padrões da Indústria e Melhores Práticas para Governança em IA:** Frameworks para o desenvolvimento responsável de IA/ML estabelecem um entendimento dos riscos potenciais do uso de IA em geral.

**Incentivo a Normas e Práticas da Indústria:** Políticos podem apoiar organizações da indústria na elaboração e promoção de frameworks responsáveis em nível da indústria, em vez de regulamentações específicas. Isso reduz o risco de regras específicas do setor onerosas, enquanto promove políticas eficazes para avaliar e controlar riscos específicos da IA em diversos setores.

**Transparência e Design Centrado no Ser Humano para Construir Confiança em Sistemas de IA/ML:** Compromissos com o desenvolvimento responsável de IA/ML, incluindo transparência e design centrado no ser humano, podem aumentar a confiança entre consumidores e trabalhadores. A ênfase no design centrado no ser humano é fundamental, especialmente em casos de tomada de decisões sensíveis.