Diagnóstico e recomendações automáticas
IA em decisões sociais
- o problema de alinhamento -

Mauricio Ayala-Rincón

Maria Eduarda Carvalho Santos (Bolsista IC), Elizângela de Freitas dos Santos (Mestranda), Mehwish Arshid (Doutoranda)

Grupo de Teoria da Computação (GTC-UnB)

Departamentos de Matemática e Ciência da Computação

UnB 60

SU22

Semana Universitária UnB
29 ago - 2 set
100 anos de Darcy Ribeiro

Universidade de Brasília, 29 de Agosto de 2022

Grupo de Teoria da Computação - GTC/UnB

- Atuação em Matemática e Ciência da Computação

Alguns egressos e cooperações

-Luiz Gadelha Jr. - LNCC Petrôpolis

-Ivan Eid Tavares Araújo - Yale University

-Daniele Nantes, Flávio L. C. de Moura, Andréia B. Avelar - UnB

-Thaynara A. de Lima, Daniel L. Ventura, André L. Galdino - UFG

-Carlos Morra Scalglioti - Siemens Munique

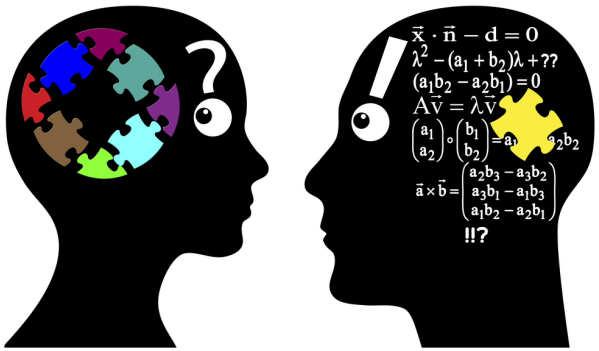
-Washington L. R. de Carvalho - IBICT, etc.

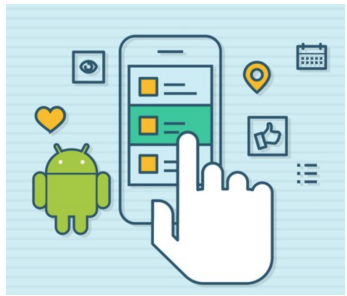
Cooperações

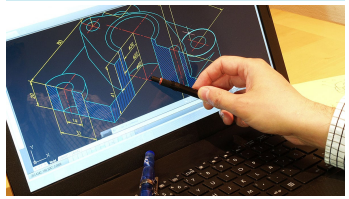
- NASA LaRC Formal Methods
- Heriot-Watt University
- Karlsruher Institute für Technologie
- King's College London
- University of Groningen
- Johannes Kepler Universität Linz

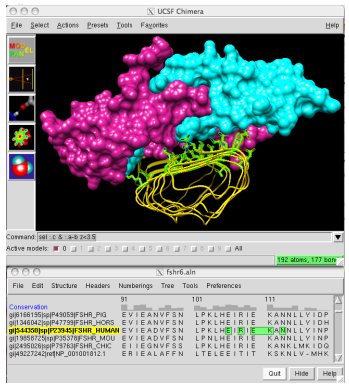
Projetos do GTC/UnB

- Lógica e Semântica de Programas de Computadores (King's College London, Johannes Kepler University Linz)



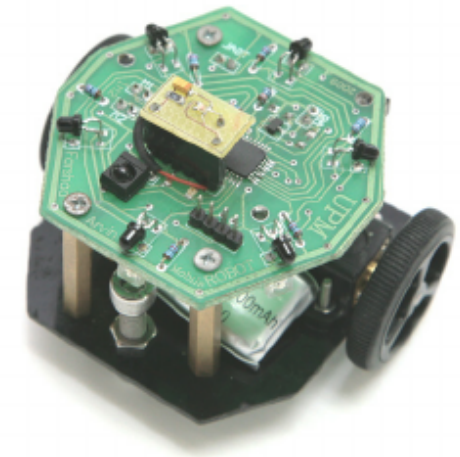


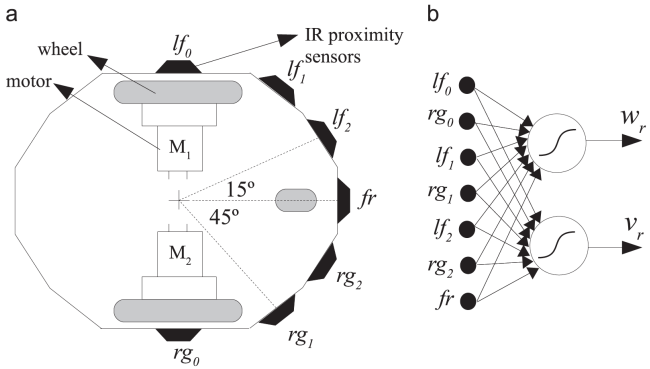


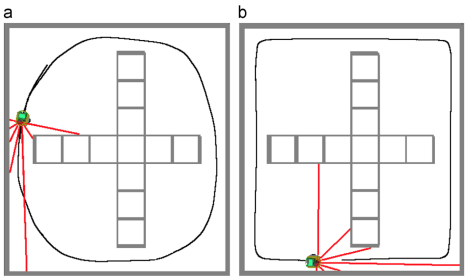


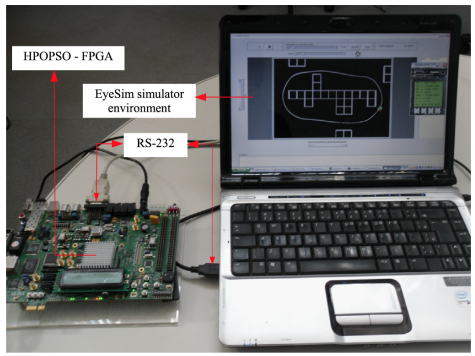
Projetos GTC/UnB

- Simulações em robôs (Cooperação Engenharia Mecatrônica)



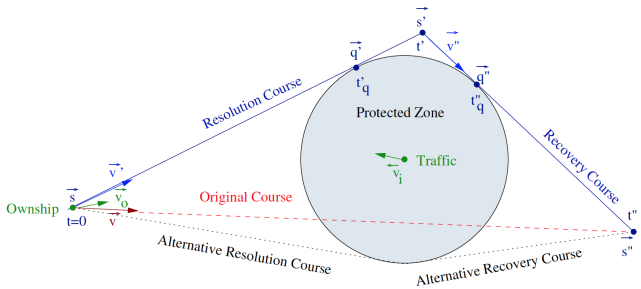


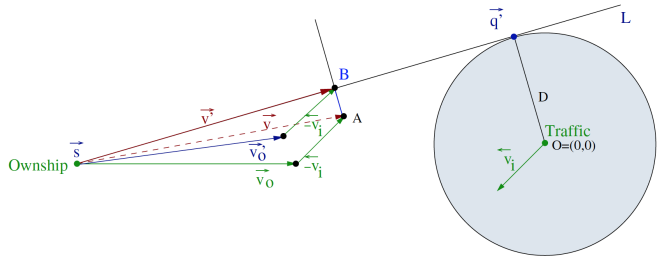


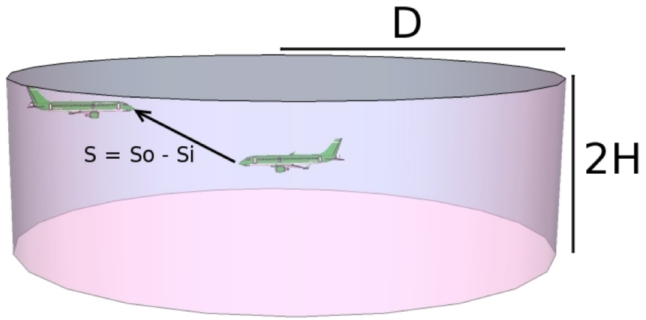


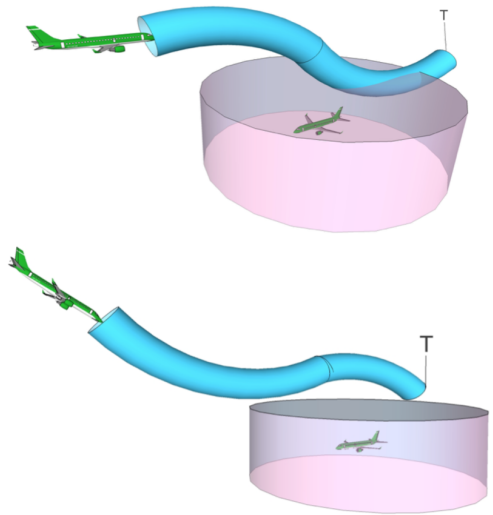
Projetos GTC/UnB

- Resolução de Conflitos em Tráfego Aéreo (cooperação NASA)









Cálculos com restrições - Somadoras e subtração

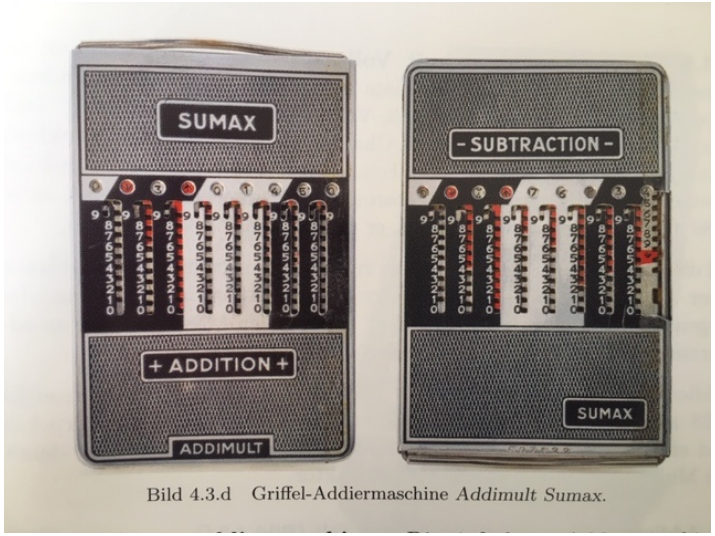
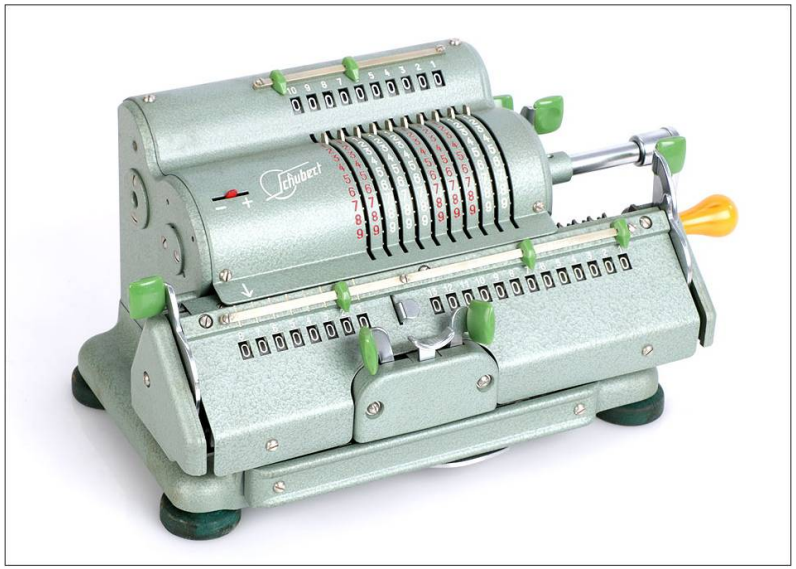


Bild 4.3.d Griffel-Addiermaschine *Addimult Sumax*.

Cálculos com restrições - Somadoras e subtração



Cálculos com restrições - Somadoras e subtração

"Acumular" 1423 e "descontar" 528

$$1423 \quad - \quad 528$$

$$1423 \quad + \quad 471 \quad = \quad 1894 \rightsquigarrow 895$$

471 é o complemento em algarismos de 528

Cálculos com restrições - Somadoras e subtração

Substração: 1752 - 1334

$$1752 \quad - \quad 1334$$

$$1752 \quad + \quad 9665 \quad = \quad 11417 \rightsquigarrow 418$$

Problemas simples que têm aplicações em Matemática Aplicada à Computação:

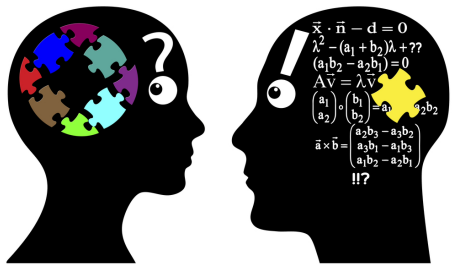
- Cálculos com restrições:
  FPGA based floating-point library for CORDIC algorithms
  DM Muñoz, DF Sanchez, CH Llanos, M Ayala-Rincón, SPL 2010.
- Verificação de Terminação de Programas:
  Formal Verification of Termination Criteria for First-Order Recursive Functions
  CA Muñoz, M Ayala-Rincón, MM Moscato, AM Dutle, AJ Narkawicz, AA Almeida, AB Avelar, TMF Ramos, ITP 2021.

Aplicação essencial:

- Deteção de erros, produção verificada de *software* e *hardware* competitivo.
- Certificação matemática de objetos computacionais.
- Valorização da industria e aplicações de IT.

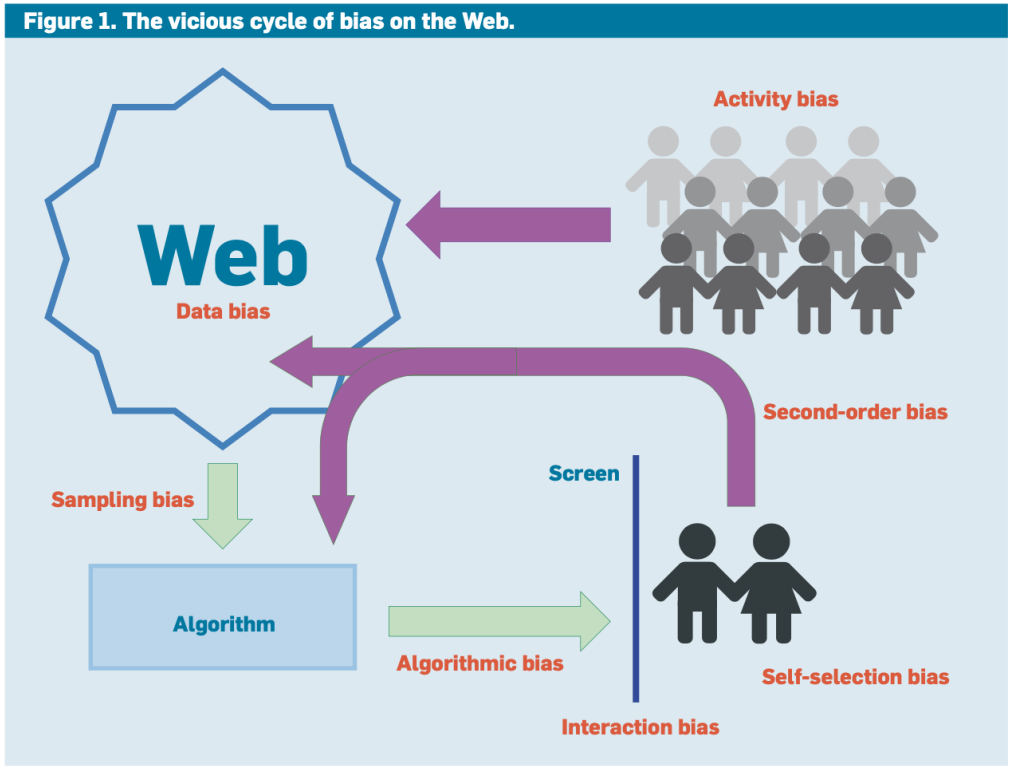
Dados

Software & Hardware



“Bias on the Web” Ricardo Baeza-Yates

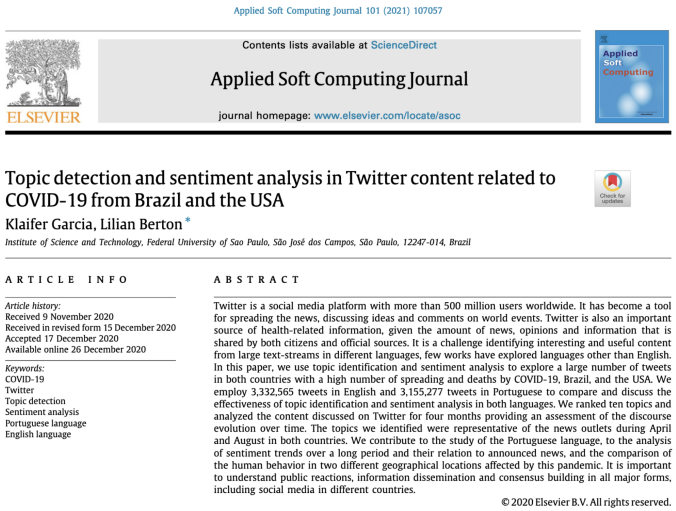
R. Baeza-Yates, *Bias on the Web* Comm. of the ACM 61(8), 2018



Discussão inicial com Maria Eduarda, Elizângela, e Mehwish:

- Viés de dados
- Viés de algoritmos

Viés de dados



Applied Soft Computing Journal 101 (2021) 107057

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA

Klaifer Garcia, Lilian Berton *

Institute of Science and Technology, Federal University of Sao Paulo, São José dos Campos, São Paulo, 12247-014, Brazil

ARTICLE INFO

Article history:
Received 9 November 2020
Received in revised form 15 December 2020
Accepted 17 December 2020
Available online 26 December 2020

Keywords:
COVID-19
Twitter
Topic detection
Sentiment analysis
Portuguese language
English language

ABSTRACT

Twitter is a social media platform with more than 500 million users worldwide. It has become a tool for spreading the news, discussing ideas and comments on world events. Twitter is also an important source of health-related information, given the amount of news, opinions and information that is shared by both citizens and official sources. It is a challenge identifying interesting and useful content from large text-streams in different languages, few works have explored languages other than English. In this paper, we use topic identification and sentiment analysis to explore a large number of tweets in both countries with a high number of spreading and deaths by COVID-19, Brazil, and the USA. We employ 3,332,565 tweets in English and 3,155,277 tweets in Portuguese to compare and discuss the effectiveness of topic identification and sentiment analysis in both languages. We ranked ten topics and analyzed the content discussed on Twitter for four months providing an assessment of the discourse evolution over time. The topics we identified were representative of the news outlets during April and August in both countries. We contribute to the study of the Portuguese language, to the analysis of sentiment trends over a long period and their relation to announced news, and the comparison of the human behavior in two different geographical locations affected by this pandemic. It is important to understand public reactions, information dissemination and consensus building in all major forms, including social media in different countries.

Discussão inicial com Maria Eduarda e Mehwish:

- Usuários de Twitter é restrito
- Análise sobre mostra elitista
- Resultados *honestos*, mas uso como recomendação para tomada de decisões ineficazes.
- Necessidade de inclusão de especialistas.

Em curso: análise de dados eleições Brasil 2022 no Twitter.

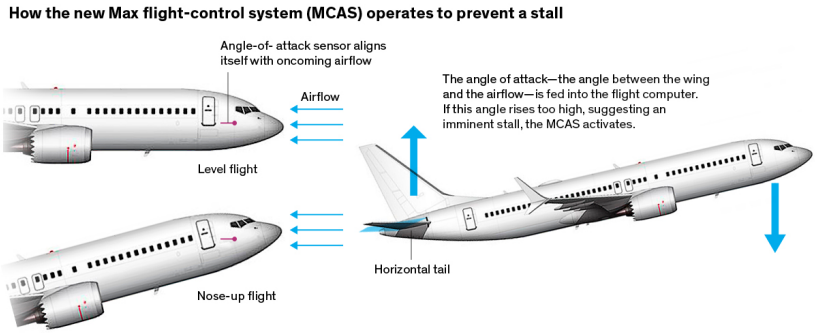
Em curso: análise de eventual viés em data e algoritmos na consulta pública do MS Brasileiro para vacinação contra Covid-19 de crianças realizado em dezembro de 2020.

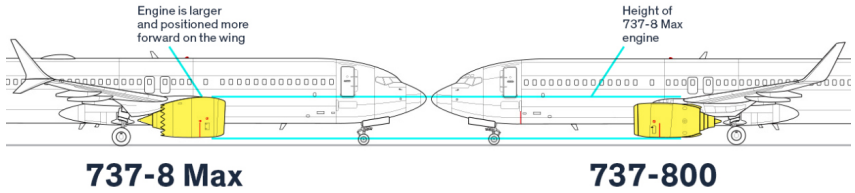
Recentes Grandes falhas em IT

"**Boeing's out-of-control autopilot**

O primeiro avião 737 Max Boeing novinho em folha, Lion Air Flight 510, caiu logo após a decolagem. Depois outro fez o mesmo. Todos a bordo morreram. Em cada caso, os pilotos tinham lutado contra um sistema de piloto automático que tomou o controle e mergulhou os aviões em sua desgraça.

Os pilotos tiveram pouco tempo para reagir a um controle de vôo, chamado MCAS, do qual pouco ou nada sabiam."





A. Regalado, *The biggest technology failures of 2019*, MIT Tech. Review, Oct. 2019

Recentes Grandes falhas em IT

"Viés de gênero no cartão de crédito da Apple

Por que um empresário rico em tecnologia obteria um limite de crédito 10 vezes maior que o de sua esposa no novo Cartão Apple, mesmo que seus bens sejam mantidos em comum? Quando alguém reclamou, um representante lhe disse: 'é apenas o algoritmo'. Um algoritmo sexista! Steve Wozniak, cofundador da Apple, disse que isso também aconteceu com sua esposa. Mas o que é o programa, e o que ele faz? A Apple e Goldman Sachs, o banco que apoia o cartão, não disseram. E esse é o problema. Existe um viés computadorizado, mas é difícil responsabilizar alguém, ou qualquer coisa.

A. Regalado, *The biggest technology failures of 2019*, MIT Tech. Review, Oct. 2019

Recentes Grandes falhas em IT

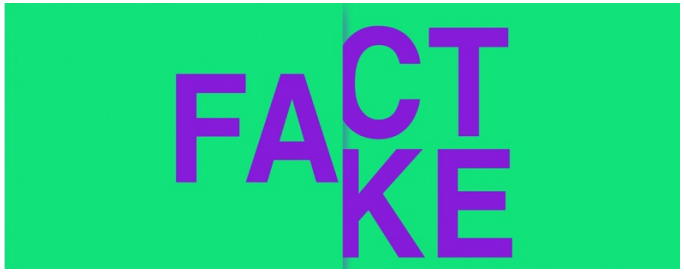
"Missed Breast Cancer Screenings Due to 'Algorithm Failure'

Quase meio milhão de mulheres idosas no Reino Unido perderam os exames de mamografia por causa de um erro de programação causado por um algoritmo de computador incorreto, e várias centenas dessas mulheres podem ter morrido precocemente como resultado.

O algoritmo errante estava no software de programação de exames de câncer de mama do Sistema Nacional de Saúde (NHS), e permaneceu sem ser descoberto por nove anos.

R. N. Charette, *450,000 Women Missed Breast Cancer Screenings Due to "Algorithm Failure"*, IEEE Spectrum, May 2018.

Desinformação versus Misinformação



"Infodemic": Usada nas epidemias de SARS (2003) e COVID-19 "fatos, misturados com medo, especulação e boatos, amplificados e retransmitidos rapidamente via tecnologias da informação"

Misinformation: "informações falsas que são divulgadas, independentemente da intenção de enganar."

Desinformation: "informações deliberadamente enganosas ou tendenciosas; narrativa ou fatos manipulados; propaganda."

Desinformação versus Misinformação

Entrevistas aos candidatos no Jornal Nacional - Agosto 2022

"A primeira vacina do mundo [contra Covid-19] foi dada em dezembro de 2020. Em janeiro, nós já estávamos vacinando no Brasil"

"O Brasil tinha 34% da sua riqueza tirada da indústria (...)"

"22% [das famílias] endividadas porque não podem pagar a conta da água, a conta de luz, a conta do gás"

"Eu prefeit@ reeleit@ com 76% dos votos"

Lupa da FSP - todas as afirmações são verdadeiras!

[1] a vacina foi a CoronaVac (Butantan-Sinovac). Em julho de 2021, seis meses após o início da campanha, apenas 13% da população brasileira tinha vacinação completa — naquela altura, 525 mil pessoas tinham perdido a vida em decorrência do novo coronavírus.
[2] IBGE (Ipeadata): em 1985 indústria de transformação foi responsável por 35,88% do PIB. Hoje, apenas 11,33%.
[3] Serasa Experian: serviços são 22,2% das contas atrasadas em julho. 67,6 milhões de inadimplentes no mês passado, maior número desde o início da série histórica do levantamento, em 2016.
[4] Tribunal Regional Eleitoral: 76,80% dos votos válidos, 36.226 votos.

Des+Misinformation - nossas responsabilidades

Table: Estimativas de Analfabetismo e Exclusão Digital em Brasil

	Analfabetismo			Exclusão		
	Absoluto	Funcional	Digital ("Readiness")	Urbana	Rural	Total
	8%	22%[1]	44° de 100 países[2]	20,6%	53,5	25,3%[3]
Referências	INAF (2018)		The Economist (2022)	Agência Brasil (2020)		

[1] É muito pouco: 17,4 milhões alcançaram o nível de proficiência em alfabetismo funcional. Analfabetos funcionais são cerca de 43,4 milhões.

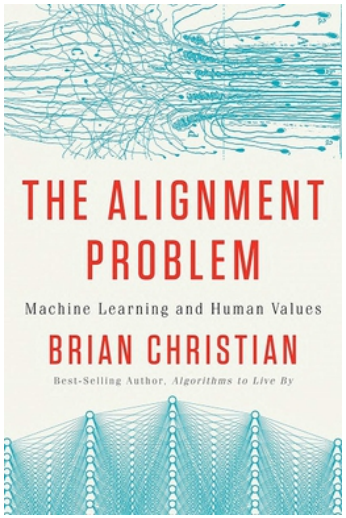
[2] "Readiness" mede aspetos como Literacia, nível de educação e preparação para usar a Internet; Confiança e Segurança, aceitação cultural da Internet; Politicas e estratégias nacionais que promovem a segurança e disseminação do uso da Internet.

[3] Um em cada quatro brasileiros não tem acesso à Internet!

Finalmente: o problema de alinhamento em IA!

Em inteligência artificial (IA), o "problema de alinhamento" refere-se aos desafios causados pelo fato de as máquinas simplesmente não terem os mesmos valores que nós. Na verdade, quando se trata de valores, em um nível fundamental, as máquinas não ficam muito mais sofisticadas do que entender que "1" é diferente de "0."

Como sociedade, estamos agora em um ponto em que começamos a permitir que as máquinas tomem decisões por nós. Então, como podemos esperar que aplicações de AI entendam que, por exemplo, eles devem decidir de uma forma que não envolvam preconceito contra pessoas de uma determinada raça, gênero ou sexualidade? Ou que a busca de velocidade, eficiência ou lucro deve respeitar a santidade última da vida humana?



AI estratificação: AI simbólica versus AI estatística

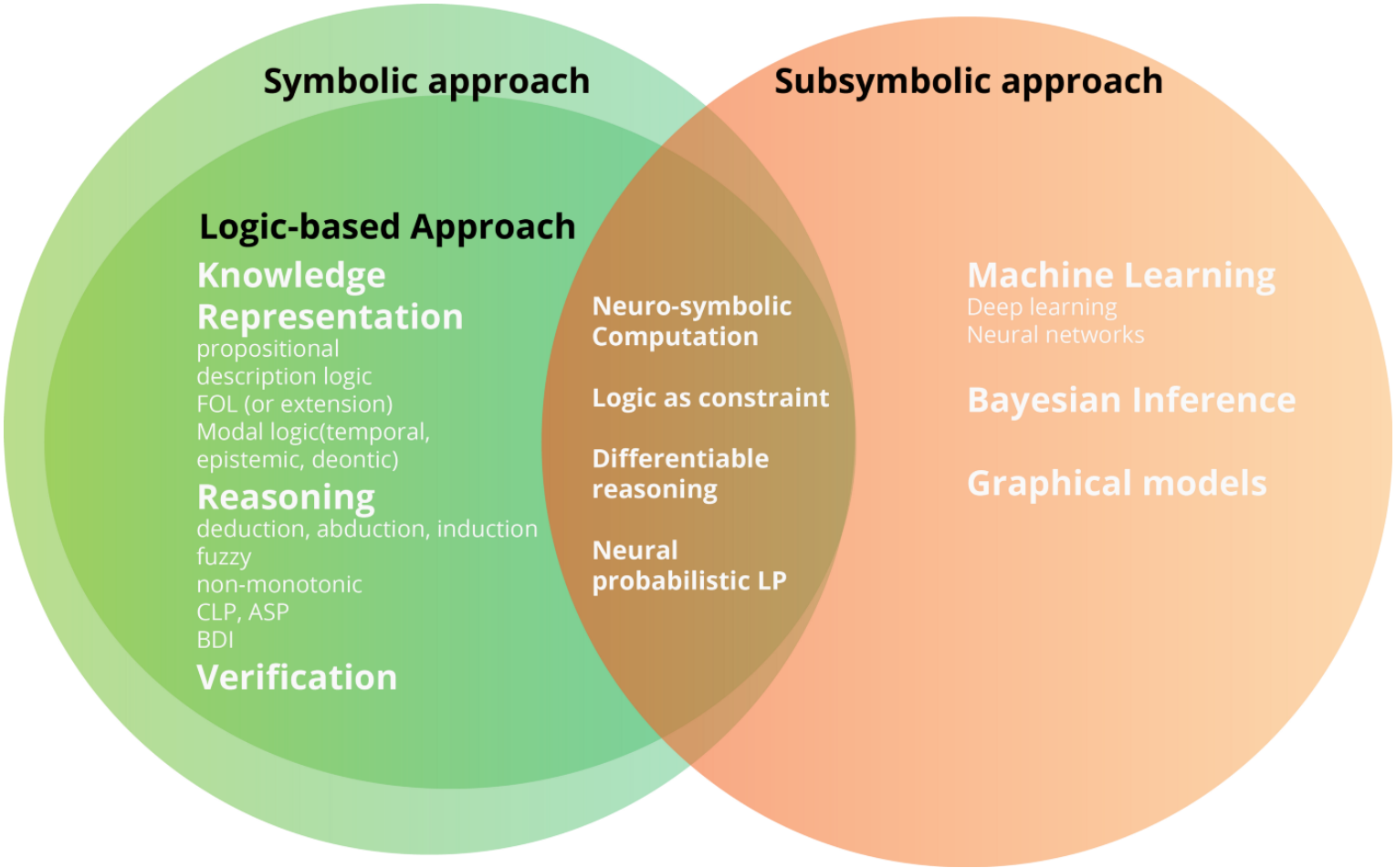


Figura de: Orhan G. Yalçın, Explainable AI, Towards Data Science, 2021.

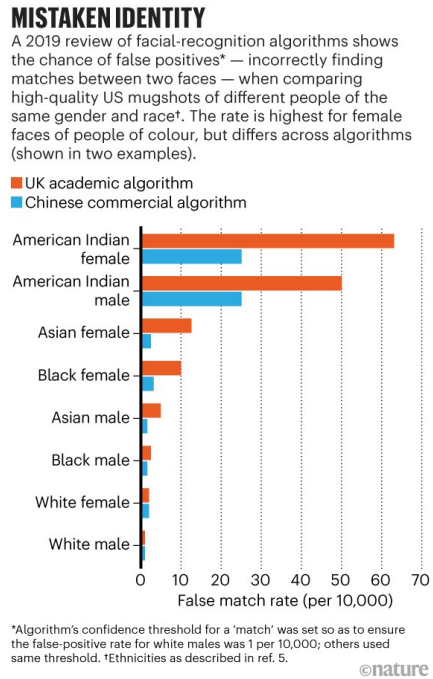
Exemplos de desalinhamento bem conhecidos

Google's image recongnition software:
classificação errada de selfies como "gorillas" detetada em 2015. Até 2018, Google não conseguiu solucionar o problema, mas simplesmente desativou ou rotulo "gorillas".

Tom Simonitte, When It Comes to Gorillas, Google Photos Remains Blind, Wired 2018

As imprecisões e problemas de privacidade do reconhecimento facial são bem conhecidos há bastante tempo e já resultaram em regulamentações de políticas especializadas em todo o mundo.

Documentário dirigido por Shalini Kantayya, 2020



Exemplos de desalinhamento bem conhecidos

g1

Maço 8, 2022 - O ministério da Economia do Brasil regulamentou a recuperação do "Dinheiro Esquecido" em contas bancárias por meio de reconhecimento facial. O sistema de reconhecimento facial deve ser usado para atualizar o nível de segurança de autenticação de usuários em contas bancárias (nível "prata" e "ouro"). Como resultado, muitos idosos e pessoas de cor foram discriminados.

> De acordo com o Ministério da Economia, "não há falhas no sistema. No entanto, diversos fatores podem estar contribuindo para esse problema, informou o orgão.
> - Como a validação da imagem ocorre por biometria facial, pode haver limitação do próprio celular do usuário;
>
> - A imagem pode estar poluída, prejudicando a identificação da pessoa. Ou seja, o usuário pode estar se fotografando com muitos objetos atrás, por exemplo. O ideal é fazer a foto com um fundo neutro."

Agosto 2022: reconhecimento facial para embarque na ponte Rio-SP

Exemplos de desalinhamento bem conhecidos

O software propietário da Northpointe COMPAS foi usado para prever quais criminosos são mais propensos a reincidir. Guiados por tais previsões, os juízes nos tribunais dos Estados Unidos (California, Florida, New York, Michigan, Wisconsin, New Mexico, Wyoming, etc) tomam decisões sobre o futuro dos réus e condenados, determinando desde valores de fiança até sentenças.

Matthias Spielkamparchive, Inspecting Algorithms for Bias, 2017

ProPublica comparou as avaliações de risco do COMPAS para mais de 10.000 presos em Broward County (Florida, 2013-2014) com a frequência com que essas pessoas realmente reincidiram: "COMPAS previu corretamente a reincidência para réus negros e brancos aproximadamente na mesma taxa". Mas quando COMPAS estava errado, foi de maneiras diferentes para negros e brancos: "Os negros são quase duas vezes mais propensos do que os brancos a serem rotulados como de maior risco, mas na verdade não reincidem". COMPAS tendia a cometer o erro oposto com os brancos: "Brancos são muito mais propensos do que os negros a serem rotulados de menor risco, mas acabam reincidindo".

FATE in AI

- Fairness: tratamento igual, honesto, sem discriminação.
- Accountability: possibilidade de atribuição de responsabilidade.
- Transparency: o resultado de um modelo de IA pode ser comunicado adequadamente.
- Explainability: capacidade de explicar como ou por que um modelo faz uma previsão.

. . . / IBM / Microsoft / Oxford / MIT / Max Plank Institute Northeastern University / CNED / . . .

Um bom exemplo no Brasil - Tô no mapa

São mais de 80 etnias indígenas presentes no bioma, além dos quilombolas, trabalhadores extrativistas, geraizeiros, vazanteiros, quebradeiras de coco, ribeirinhos, pescadores artesanais, barranqueiros, fundo e fecho de pasto, sertanejos, ciganos, entre tantos outros.

"Aplicativo desenvolvido para que povos, comunidades tradicionais e agricultores familiares brasileiros realizem o automapeamento de seus territórios."
Instituto de Pesquisa Ambiental da Amazônia (IPAM), Instituto Sociedade, População e Natureza (ISPN), e Rede Cerrado.

- Aplicação alinhada com a realidade de populações tradicionais do Cerrado: mapeamento via GPS e subsequente fornecimento de dados verificados pelas comunidades via internet.
- Participação de times multidisciplinares.

Recomendações típicas para alinhamento em AI

1. **Enriquecimento com avaliação humana** - Avaliação do impacto de previsões incorretas e, quando razoável, projeção de sistemas com processos de revisão "human-in-the-loop".
2. **Avaliação do viés de data e computacional** - Elaboração de processos que permitam compreender, documentar e monitorar preconceitos.
3. **Explicabilidade e transparência** - Desenvolvimento de ferramentas e processos para melhorar continuamente a transparência e a explicabilidade dos modelos de aprendizagem de máquinas onde for razoável. Mesmo que em certas situações a precisão possa diminuir, os ganhos em transparência e explicabilidade podem ser significativos.
4. **Operações reprodutíveis** - Desenvolvimento da infra-estrutura necessária para permitir um nível razoável de reprodutibilidade.
5. **Estratégia de deslocamento** - Identificação e documentação de informações para que os processos de mudança comercial possam ser desenvolvidos para mitigar o impacto para os trabalhadores que estão sendo automatizados.
6. **Precisão** - Desenvolvimento de processos para garantir precisão e funções métricas de custo alinhadas às aplicações específicas do domínio.
7. **Confiança e privacidade** - Construção e comunicação de processos que protejam e tratem os dados com as partes interessadas que possam interagir com o sistema direta e/ou indiretamente.
8. **Conscientização do risco dos dados** - Desenvolvimento e melhoria de processos e infra-estrutura razoáveis para garantir que os dados e modelos de segurança estejam sendo levados em consideração. Sistemas autônomos de tomada de decisão abrem as portas para novas brechas potenciais de segurança.

Sucesso =
Alinhamento e
Multidisciplinaridade!

Bem-vindos ao Instituto de Ciências Exatas!

https://www.mat.unb.br/pg/45/Teoria-da-Computacao

Universidade de Brasília
Instituto de Ciências Exatas

Departamento de Matemática

Português

Buscar

Home Institucional Graduação Pós Graduação Pesquisa Extensão Pessoas Contato

- Notícias
- Seminários
- Concursos
- Eventos
- Links e Formulários
- Mídia MAT
- Galeria
- Comissões

Teoria da Computação

A pesquisa está focada no desenvolvimento de estruturas matemáticas e formais para dedução e computação. Especificamente, arcabouços lógicos como os sistemas de reescrita, o cálculo Lambda, as substituições explícitas, e os sistemas nominais são estudados e suas aplicações em computação e dedução investigadas.

Colaboradores do grupo includem coautores brasileiros e estrangeiros: César Muñoz, Maribel Fernández, Fairouz Kamareddine, Flávio L.C. de Moura, Daniel Ventura, entre outros.

Linhas de Pesquisa

- Teoria da reescrita
- Teoria de tipos
- Lógica formal e computacional
- Teoria de prova
- Dedução formal e equacional

Atividades

Página do seminário de Teoria da Computação | Eventos | Publicações

Quem Somos

Daniele Nantes Sobrinho
Aplicações de Estruturas Formais em Dedução Equacional e Modelos Computacionais.

Orientadora de mestrado

Maurício Ayala Rincon
Aplicações das Teorias de Reescrita, Tipos e Prova em Formalização e Dedução.

Orientador de mestrado e doutorado

Bem-vindos ao Instituto de Ciências Exatas!

https://www.mat.unb.br/∼ayala

Mauricio Ayala Rincón, Dr. rer. nat.
Professor Titular
Teoria da Computação
Departamentos de Ciência da Computação e Matemática
Universidade de Brasília

Endereço:
Departamento de Matemática, Universidade de Brasília
Campus Universitário Darcy Ribeiro, Asa Norte
70910-900 Brasilia D. F., Brasil
Tels. +55-61-3307 2441|2442| +55- 61-3107 6453 | 3676 Fax +55-61-3273 2737
e-mail: *ayala[at]unb.br*

Publicações · Cursos <<<== Inicio 17 Agosto - atividades remotas 2020-1
PVS Class 2017 (associado a ITP 2017) · PVS Tutorial for Mathematicians (associado a SW in Math 2020
Atividades profissionais · CV Lattes
Grupo de Teoria da Computação

Tópicos de pesquisa:
Propriedades e aplicações dos sistemas de reescrita de termos e suas extensões. *Links* relacionados
TRS PVS teoria de reescrita · Alg. evolut. para ordenação de permutações

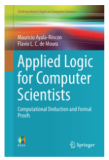
---

English Español

Oportunidades

- Bolsas de doutorado Edital de Seleção com inscrição até 18 de Agosto -estendido para 1ro de Setembro, 2020. Interessados no tema 2: *Algorítmica e Teoria de Sequenciamento de Informação Genômica*, entrar em contato.
- RTA (1983 ... 2015) and TLCA (1993 ... 2015) evolved to Int. Conf. on Formal Structures for Computation and Deduction FSCD (2016 ... 2019, 2020), FSCD 2021 em Buenos Aires (a 12/2/2021, p 15/2/2021) cfp.
- 30th Int. Symp. on Logic-based Program Synthesis and Transformation LOPSTR 2020, Bologna, 7-9 Setembro, 2020.
- 15th Int. Logical and Semantic Frameworks, with Applications LSFA 2020, 26-28 Agosto, 2020.
- Conferencias em curso no GTC/UnB

M. Ayala-Rincón & Flávio L.C. de Moura, *Fundamentos da Programação Lógica e Funcional - O Princípio de Resolução e a Teoria de Reescrita* -, Course Notes, Ed. UnB, December 2014. Em Portugês.



M. Ayala-Rincón & Flávio L.C. de Moura, *Applied Logic for Computer Scientists: Computational Deduction and Formal Proofs*, Springer, 2017.

---

ayala[at]unb.br