

# Inteligência Artificial e Aprendizado de Máquina: Estado atual, tendências e aspectos sociais

André C. P. L. F. de Carvalho





# Tópicos

---

- Inteligência Artificial
- Ciência de Dados
- Aprendizado de Máquina
- Aspectos sociais
- Decisões baseadas em evidências
- Conclusão



# CEPID CeMEAI

- Centro de Ciências Matemáticas Aplicadas à Indústria
  - ICMC-USP
- Início em 2013, duração de até 11 anos
  - Orçamento para primeiros 5 anos de 15 milhões de Reais da FAPESP
    - Mesma quantidade obtida de empresas
- Objetivo principal
  - Transferir conhecimento em computação, matemática e estatística para empresas



# CEPID CeMEAI



Português

Search

## Center for Research in Mathematical Sciences Applied to Industry



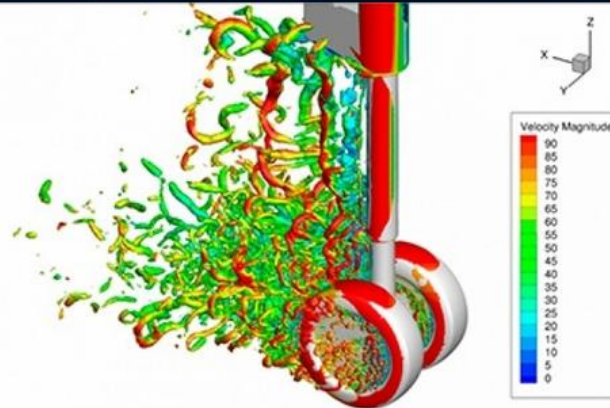
### RIDC

Research, Innovation and Dissemination Centers

### About the Program

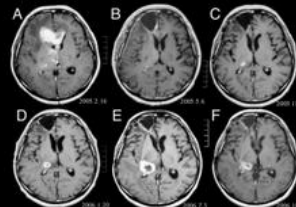
#### RIDCs

- Center for Metropolitan Studies
- Brazilian Research Institute for Neuroscience and Neurotechnology
- Center for Cell-Based Therapy
- Center for Computational Engineering and Sciences
- Center for Research and Development of Functional Materials
- Center for Research and Innovation in Biodiversity and Drug Discovery
- Center for Research in Mathematical Sciences Applied to Industry
- Center for Research on Inflammatory Diseases
- Center for Research on Redox Processes in Biomedicine
- Center for Research on Toxins, Immune-Response and Cell Signaling
- Center for Research, Teaching, and Innovation in Glass
- Center for the Study of Violence



### Researchers analyze aircraft landing gear noise

A group at a Brazilian university, partnering with engineers from Boeing, are using computer simulations to identify sources of noise produced by aircraft undercarriages



### CEPID 2013-2024

About the Center for Research in Mathematical Sciences Applied to Industry (CeMEAI)

WEBSITE: [www.cemeai.icmc.usp.br](http://www.cemeai.icmc.usp.br)

Principal Investigator: José Alberto Cuminato

Education and Knowledge Diffusion Coordinator: Lucio Tunes dos Santos

Technology Transfer Coordinator: Francisco Louzada Neto

#### HOST INSTITUTION

University of São Paulo (USP)  
Av. do Trabalhador São-Carlense, 400  
Caixa Postal 688  
13566-590 - São Carlos, SP - Brasil  
Tel.: (16) 3373-8159  
[contatocepid@icmc.usp.br](mailto:contatocepid@icmc.usp.br)  
[jacumina@icmc.usp.br](mailto:jacumina@icmc.usp.br)

#### ASSOCIATED INSTITUTIONS

Institute of Aeronautics and Space (IAE)  
Federal University of São Carlos (UFSCAR)  
São Paulo State University (UNESP)  
State University of Campinas (UNICAMP)

#### RESEARCHERS

Research projects supported in this Center at FAPESP's Virtual Library

Center leaflet (pdf):



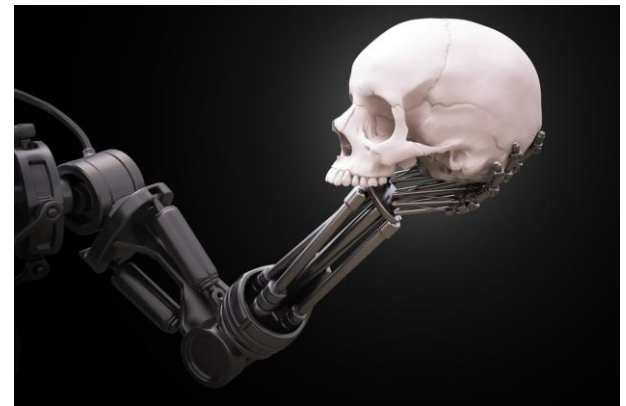
<http://cepid.fapesp.br/en/centro/15/>

# O que é Inteligência Artificial?

*"É o estudo de como computadores podem fazer tarefas de uma forma melhor do que as pessoas atualmente fazem."*

Elaine Rich, 1990

- Inteligência externa a um ser vivo
  - Inteligência de máquina
  - Inteligência simulada
  - Máquinas que pensam
    - Alan Turing, 1950





# Apogeu da IA

---

- Automação crescente
- Principal protagonista:
  - Aprendizado de Máquina (AM)
- Protagonismo de AM é influenciado por
  - Big Data
  - Ciência de Dados



# Revolução do conhecimento<sup>2</sup>

---

- Países disputam protagonismo
  - Aumentam investimentos
  - Incentivam adoção por indústrias
  - Avaliam efeitos
  - Criam regulações
  - União Europeia
    - Reino Unido



# Reino Unido

---

- Primeira Ministra, Theresa May
  - Defendeu em Davos que Reino Unido lidere o mundo no uso seguro e ético de IA
  - Criou Center for Data Ethics and Innovation
    - Órgão assessor para garantir uso ético e seguro de tecnologias baseadas em dados
    - Foco do Oxford University's Future of Humanity Institute e do Alan Turing Institute





# Reino Unido

---

- Orçamento será usado para veículos autônomos, para torná-los comum até 2021
- Desde 2015, empresas de IA têm sido criadas toda semana
- Sediou criação do Alexa da Amazon e do DeepMind AI, da Google



# Reino Unido

---

- Comitê da Câmara dos Lordes
  - Reino Unido deve ser um líder mundial no desenvolvimento de Inteligência Artificial
    - Melhor forma de fazer isso é atribuir a Ética papel central de desenvolvimento e uso de Inteligência Artificial
      - Movimento deve ser internacionalmente liderado pelo Reino Unido
      - Ao invés de aceitar passivamente suas consequências



# Reino Unido

---

- Alan Turing Institute
  - Centro nacional de pesquisa para ciência de dados e inteligência artificial
    - Criado em 2015
  - Cientistas das principais Universidades Britânicas
  - Realiza pesquisas em alguns dos maiores desafios em ciência, sociedade e economia
    - Grupos de pesquisa em Ética e Transparência



# Passado

---

**Poucos dados**



# Presente

## Inundação de dados



<https://inside.igneous.io/storing-files-beyond-big-data>



# Inundação de dados

---

- Dados gerados por empresas
  - De 2016 a 2020, é estimado que a geração de dados vai aumentar em 4.300%
    - Dados produzidos internamente
  - Atualmente, empresas usam apenas uma pequena parcela dos dados que produzem e armazenam
    - Empresas já sofrem para lidar com esses dados
- Além dos dados internos, haverá um grande aumento na quantidade de dados externos

# Causas da inundação

- Avanços recentes nas tecnologias para
  - **Aquisição, armazenamento, transmissão e processamento** de dados
  - Maior **quantidade**, mais **rapidez** e menor **custo**





# Causas da inundação

- Avanços recentes nas tecnologias para
  - **Aquisição, armazenamento, transmissão e processamento** de dados
  - Maior **quantidade**, mais **rapidez** e menor **custo**

**Big Data**







# Neste século

---

- Mais conjuntos de dados
  - Maiores, mais complexos e dinâmicos
  - Big Data
- Novos desafios de análise de dados
- Necessidade de desenvolvimento de novas ferramentas
  - Motivaram avanços na análise de dados
    - Ciência de dados

# Ciência de Dados

- Estuda princípios e técnicas para extrair conhecimento de um conjunto de dados
  - Novo, relevante e útil
- Pergunta chave da área:
  - Como extrair (de forma eficiente) conhecimentos em (grandes) conjuntos (fluxos) de dados





# Ciência de Dados

---

- Técnicas tradicionais de análise de dados permitem apenas consultas simples
  - Quantos itens de um produto em particular foram vendidos em um dado dia?
  - Não conseguem responder consultas do tipo:
    - Que novo filme eu gostaria de assistir?
    - Dado o que estou sentindo, posso estar doente?
    - O que significa esse texto em chinês?
- Por isso usamos **Aprendizado de Máquina**

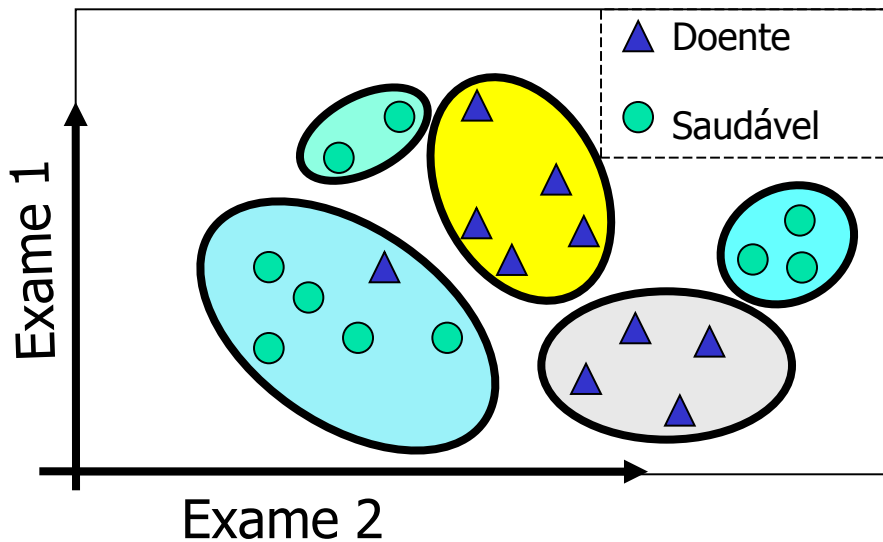


# Aprendizado de Máquina

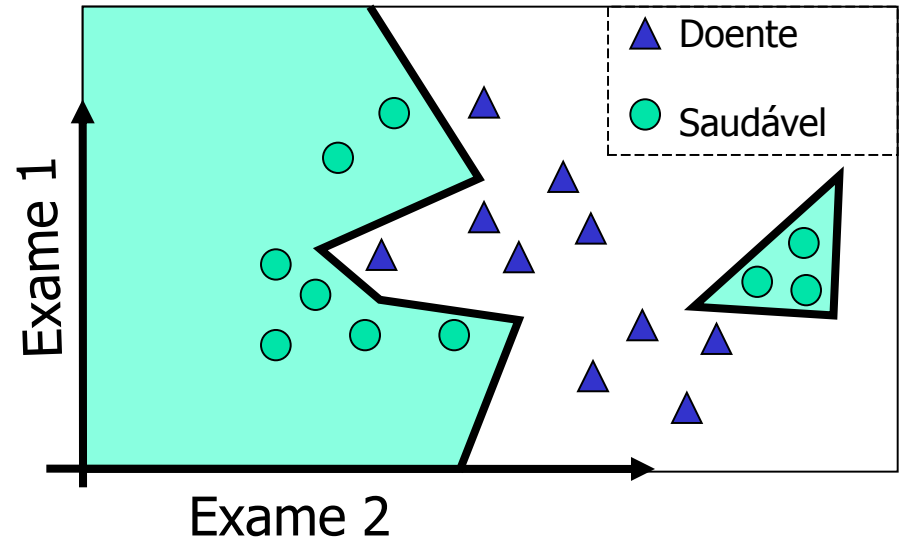
---

- Investiga técnicas capazes de aprender a resolver problemas
  - De forma automática, reduzindo (eliminando) intervenção humana
- Bem sucedido em vários problemas reais de modelagem
  - Descritivos
  - Preditivos

# Aprendizado de Máquina



Descritiva  
Agrupamento



Preditiva  
Classificação

# Quando começou?

- Filósofo grego Aristóteles criou a lógica moderna (384-322 AC)
  - Conjunto de seis textos sobre lógica (*organon*: instrumento)
  - Reconhecia duas formas de argumentação
    - Dedução (raciocínio dedutivo)
      - A partir de um ou mais termos gerais chega-se a uma conclusão particular logicamente correta
    - Indução (raciocínio indutivo)
      - Generaliza a partir de conhecimentos particulares
      - **Aprendizado**



Lógica  
Top-down

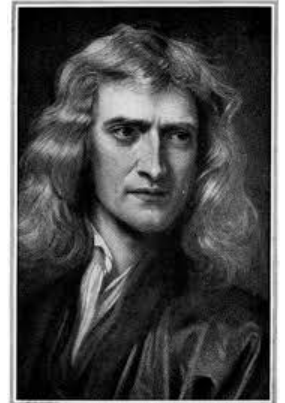
Lógica  
Bottom-up



# Um dos marcos

---

- Isaac Newton publica *Principia* em 1687
  - Princípios Matemáticos da Filosofia Natural
    - Conjunto de 3 livros
      - 3 leis de movimento
        - Lei da inércia
        - Princípio fundamental da dinâmica
        - Lei da ação e reação

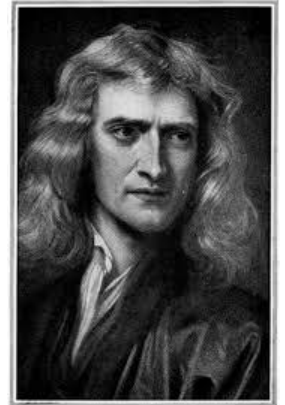




# Um dos marcos

---

- Isaac Newton publica *Principia* em 1687
  - Princípios Matemáticos da Filosofia Natural
    - Conjunto de 3 livros
      - 3 leis de movimento
        - Lei da inércia
        - Princípio fundamental da dinâmica
        - Lei da ação e reação
      - 4 regras de indução
        - 3a regra: O que é verdade para tudo que vimos é verdade para tudo no Universo



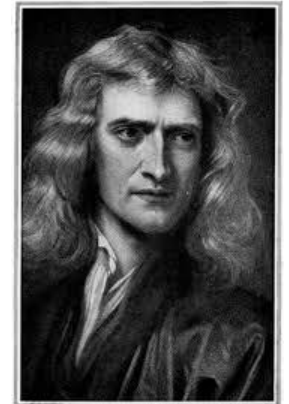




# Um dos marcos

---

- Isaac Newton publica *Principia* em 1687
  - Princípios Matemáticos da Filosofia Natural
    - Conjunto de 3 livros
      - 3 leis de movimento
        - Lei da inércia
        - Princípio fundamental da dinâmica
        - Lei da ação e reação
      - 4 regras de indução
        - 3a regra: O que é verdade para tudo que vimos é verdade para tudo no Universo
          - Indução: generaliza a partir de conhecimentos particulares





# Aplicações de AM

---

- Extração de conhecimento de dados requer o apoio de especialistas para várias tarefas
  - Construção e manutenção do conjunto de dados
  - Exploração dos dados
  - Escolha de técnicas para melhoria de qualidade e pré-processamento dos dados
  - Seleção de algoritmos de modelagem
  - Ajuste de hiperparâmetros dos algoritmos
  - Como melhor analisar e validar os resultados
  - Como manter o modelo válido ao longo do tempo



# Aplicações de AM



---

- Extração de conhecimento de dados requer o apoio de especialistas para várias tarefas
- Construção e manutenção do conjunto de dados
  - Exploração dos dados
  - Escolha de técnicas para melhoria de qualidade e pré-processamento dos dados
  - Seleção de algoritmos de modelagem
  - Ajuste de hiperparâmetros dos algoritmos
  - Como melhor analisar e validar os resultados
  - Como manter o modelo válido ao longo do tempo



# Aplicações de AM

---

- Grande número de opções 
  - Várias com hiperparâmetros
  - Impossível testar todas as combinações
- Cada algoritmo tem um viés 
  - Que o torna mais adequado para uma dada conformação dos dados
    - Influenciado pelos valores dos hiperparâmetros
  - Permite reduzir espaço de busca



# AutoML

---

- Automatizar AM (do início ao fim)
  - Aprendizado do aprendizado
  - Auto-Weka, Auto-sklearn, TPOT
  - MIT, University of Freiburg, USP
  - IBM PredictiveInsight, IBM SPSS Modeler , Facebook, Google, Microsoft
- Apoio ao especialista e usuário leigo
  - Algumas tarefas podem ser automatizadas
    - Meta-aprendizado

# Aprendizado em fluxos de dados

- Problemas reais são dinâmicos
  - Dados são gerados continuamente e em alta velocidade
  - Médio > grande > extremamente grande
  - Fluxos de dados
  - Novos algoritmos de AM e modificação de algoritmos existentes



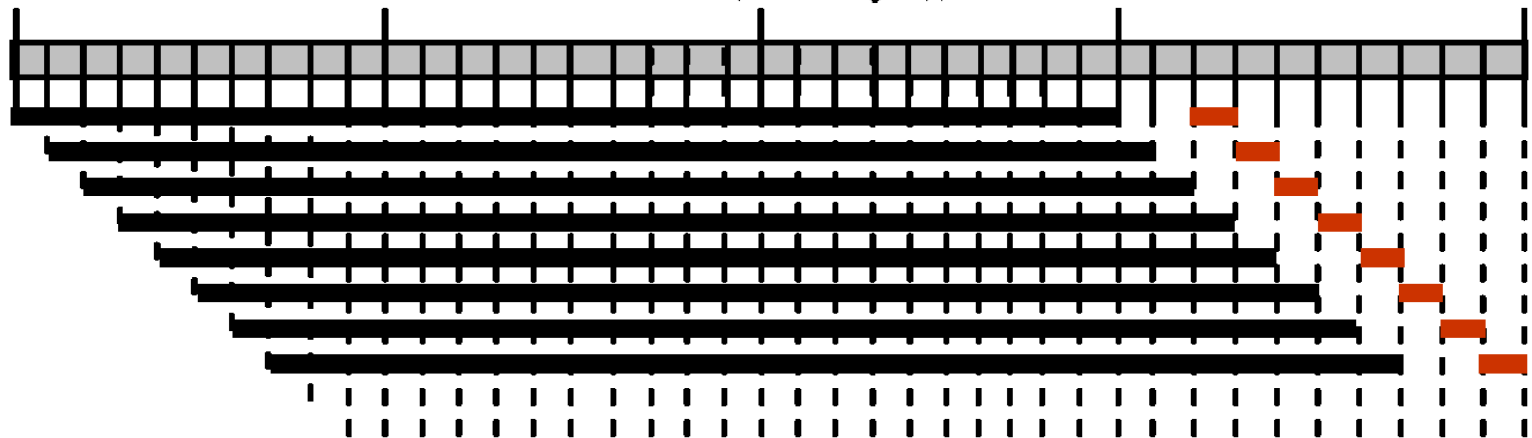
# Previsão de tempo de viagem

- Previsão de tempo de viagem 3 dias antes
  - Redefinir veículos e motoristas

USP

U.PORTO

Data (in days)



Training set  
Test set



# Revolução Industrial

---

*Não existe ação mais legítima de um legislador do que proteger aqueles cujas vidas serão prejudicadas pelas máquinas*

John Stuart Mill (1800s)





# Preocupação social

---

- Propósito de avanços tecnológicos:
  - Melhora na qualidade de vida
- Novas leis devem ser criadas
  - Proteger indivíduos
    - Prevenir e punir uso abusivo
- Mudanças na educação
- Avanços
  - Engajamento social crescente
  - Decisões baseadas em evidências



# Mudança no mercado de trabalho

---

- EUA, Século XIX
  - Quantidade de tecido que um tecelão era capaz de produzir em 1 hora aumentou em 50x
    - Número de operários necessários para produzir um tamanho de tecido diminuiu em 98%
    - Reduziu custo de tecidos, e aumentou demanda
    - Criou 4x mais empregos no longo prazo
  - Surgimento dos automóveis reduziu empregos relacionados a cavalos
    - Novos empregos, não só na indústria automobilística



# Mudança no mercado de trabalho

---

- Algumas previsões:
  - Mercado de trabalho mudará cada vez mais rapidamente
    - Estudantes atuais terão 10 a 14 empregos até os 38 anos
    - 10 profissões de maior demanda em 2010 não existiam em 2004
    - 65% das crianças que ingressam na escola primária hoje trabalharão em profissões que ainda não existem
  - Expectativa de vida de quem nasce hoje é de 100 anos

# AM e mercado de trabalho

- Como toda nova tecnologia, mercado de trabalho será afetado
- Algumas profissões são mais vulneráveis a automação
  - Maior chance de desaparecer
  - Novas oportunidades

## Catalogue of fears

Probability of computerisation of different occupations, 2013  
(1 = certain)

| Job                         | Probability |
|-----------------------------|-------------|
| Recreational therapists     | 0.003       |
| Dentists                    | 0.004       |
| Athletic trainers           | 0.007       |
| Clergy                      | 0.008       |
| Chemical engineers          | 0.02        |
| Editors                     | 0.06        |
| Firefighters                | 0.17        |
| Actors                      | 0.37        |
| Health technologists        | 0.40        |
| Economists                  | 0.43        |
| Commercial pilots           | 0.55        |
| Machinists                  | 0.65        |
| Word processors and typists | 0.81        |
| Real-estate sales agents    | 0.86        |
| Technical writers           | 0.89        |
| Retail salespeople          | 0.92        |
| Accountants and auditors    | 0.94        |
| Telemarketers               | 0.99        |

Source: "The Future of Employment: How Susceptible are Jobs to Computerisation?", by C. Frey and M. Osborne (2013)

Economist.com



# AM e mercado de trabalho

---

- AM para contratação de pessoal
  - Mercado de US\$ 100 bilhões, com quase 100 startups
  - Produtos
    - Prediz se perfil psicológico de candidato se encaixa na cultura da empresa
    - Prediz profissionais que podem mudar de emprego (maior chance de aceitar oferta)
    - Prediz desempenho de candidato no trabalho a partir de atributos pessoais



# AM e mercado de trabalho

---

- AM identifica traços que podem ser relevantes para bom desempenho
  - Habilidades cognitivas
  - Humildade intelectual
  - Capacidade de aprender
  - Mais importantes que formação em boas escolas



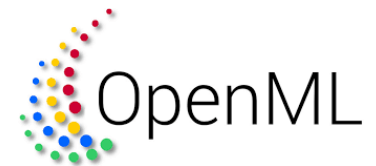
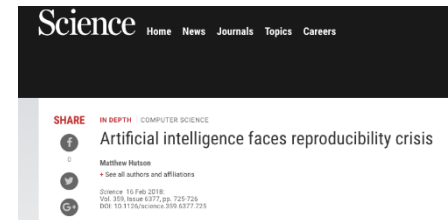
# AM e mercado de trabalho

---

- Mudando como candidatos são selecionados
  - Diplomas estão perdendo importância, o que importa é conhecimento
  - Persistência e dedicação estão sendo considerados mais importantes que QI
  - Atividades extracurriculares estão se tornando tão importantes quanto experiência
  - Sucesso em antiga empresa não implica em sucesso na nova
  - Fotos com bebida em redes sociais tem menos relação com baixo desempenho que comentários postados

# Reprodutibilidade e Transparência

- Reprodutibilidade
  - Disponibilização e curadoria de dados e códigos
- Privacidade
  - *Fair Information Practices* (FIPs) for data
    - Direito a privacidade
- Transparência
  - *General Data Protection Regulation (GDPR-EU)*
    - Direito a informação







# Direito à explicação

---

- AM pode gerar modelos “caixa preta”
- Abril de 2016 parlamento europeu aprova a *General Data Protection Regulation* (GDPR)
  - Regulação geral de proteção de dados
  - Regula coleta, armazenamento e uso de informação pessoal
  - Artigo 22: *Automated individual decision-making, including profiling*
    - Tomada de decisão individual automatizada, incluindo criação de perfis



# Proteção de privacidade

---

- *Fair Information Practices* (FIPs) for data
- Conjunto de 10 princípios para:
  - Coleta de dados
  - Acesso aos dados
  - Uso dos dados
  - Compartilhamento de dados



# Ciência de Dados para o Bem

---

- Traz benefícios sociais para pessoas e comunidades
  - Bons serviços de saúde para todos
  - Desenvolvimento econômico de países pobres
  - Educação pública de qualidade
  - Energia limpa e barata
  - Melhor exercício da cidadania
  - Proteção ambiental
  - Meios de transportes mais seguros, rápidos e limpos



# Decisões baseadas em evidências

---

- Movimento para que decisões sejam mais efetivas
  - Tragam mais benefícios
  - Reduzir influência de crenças e dogmas
  - Permitir políticas de estado
  - INGSA
    - Governo com ciência
    - Diplomacia científica

- International Network for Science Advice
  - *Science Government*
  - Estimula participação da ciência nas decisões publicas
    - Uso de evidências científicas
  - Capacitação de interlocutores governo-academia





# Pesquisas no ICMC-USP

---

- **Bioinformática e Medicina**
  - Diagnóstico de câncer
  - Diagnóstico de demência
  - Prevenção de acidentes com idosos
    - Leiden University Medical Center)
- **Biometria e segurança**
  - Sequência de digitação e padrão de caminhada
  - DéjàVu, recontando a estória
    - Polícia Federal e DARPA e UNICAMP



# Pesquisas no ICMC-USP

---

- Ciência de Materiais
  - Projeto de novos materiais
    - UFSCar, Shell
- Finanças
  - Sistema automático de cobranças
    - Virgos IP *Solutions*
  - Detecção de fraudes
    - NEC (Alemanha)



# Pesquisas no ICMC-USP

---

- Processos industriais
  - Qualificação de madeiras
    - Sguario Indústria de Madeira, Faber Castell e Universidade de Oulu (Finlândia)
  - Otimização de alocação de peças cerâmicas em fornos industriais
    - *Optimized Decision Making* – ODM – PIPE FAPESP
- Meio ambiente
  - Modelagem de distribuição de espécies
    - Instituto Florestal





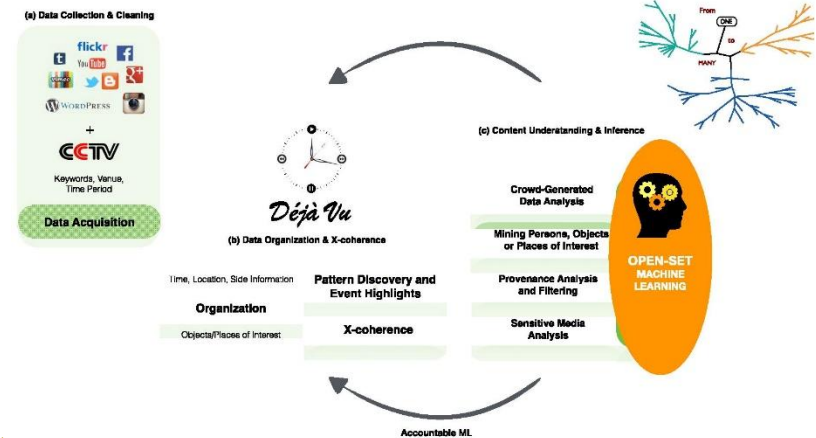
# Pesquisas no ICMC-USP

---

- Órgãos públicos
  - Recuperação de textos em livros antigos
    - Biblioteca Nacional da França Universidades de Bordeaux e Universidade de La Rochelle (França)
  - Monitoramento de pacientes em UTIs
    - Hospital São João, Universidade do Porto (Portugal)
  - Análise contínua de dados de obstetrícia
    - Compartilhamento de dados de hospitais portugueses
    - Detecção de anomalias
    - Melhora na recomendação de partos normais e por cesariana

# Projeto DejaVu

- DéjàVu: Feature-Space-Time Coherence from Heterogeneous Data for Media Integrity Analytics and Interpretation of Events



Finding arrow of time

# Projeto DeJaVu



## DéjàVu

**Feature-Space-Time Coherence from Heterogeneous Data  
for Media Integrity Analytics and Interpretation of Events**

**Anderson Rocha (Associate Professor)**

*Microsoft Research Faculty Fellow  
Google Faculty Research Awardee  
Tan Chin Tuan Fellow  
IEEE Senior Member  
[anderson.rocha@ic.unicamp.br](mailto:anderson.rocha@ic.unicamp.br)*

**Reasoning for Complex Data (RECOD) Lab.**

*Institute of Computing,  
University of Campinas (Unicamp)*

*Av. Albert Einstein, 1251 – Cidade Universitária  
CEP 13083-970 • Campinas/SP – Brasil*



Image Credits (L to R): Peter Michael Hamel; ABC News; CNN; Child Rescue Coalition; thesun.co.uk  
Background Music: Anthony Greninger - A Heavy Heart; Ludovico Einaudi - Primavera





# Tendências futuras

---

- Ciência de Dados a toda hora e em todo lugar
  - Automatização de Ciência de Dados
  - Detecção de novidades
- Uso crescente de técnicas de Ciência de Dados
  - Benefícios para população e especialista
  - Qualidade de vida
- Disponibilidade de dados, transparência de dados e modelos e privacidade



# Tendências futuras

---

- Aplicações
  - Análise de sentimentos para tomada de decisão
  - Benefícios sociais
  - Novos materiais
  - Personalização de ensino, saúde, lazer e investimentos
  - Robôs domésticos
  - Veículos autônomos



# Conclusão

---

- Inteligência Artificial
- Ciência de Dados
- Bem estar e qualidade de vida
- Decisões baseadas em evidências
- Tendências
- Conclusão



# Agradecimentos

---

**IBM Research | Brasil**

