Data Science

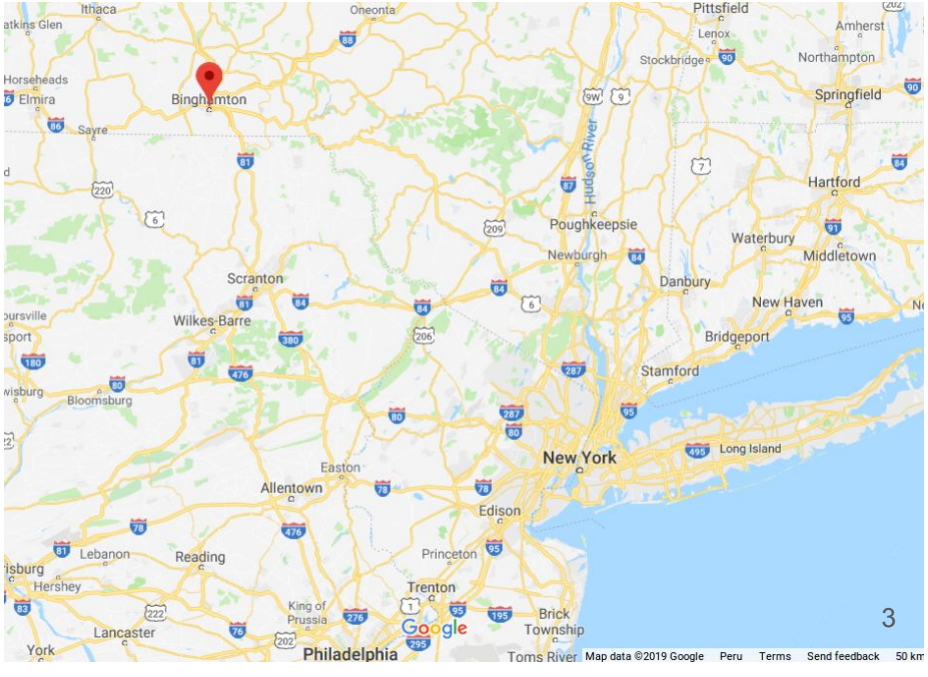
WWW.DATASCIENCE.PE

Estado de la investigación en Inteligencia Artificial

Gissella Bejarano
PhD student Binghamton University

Acerca de mí

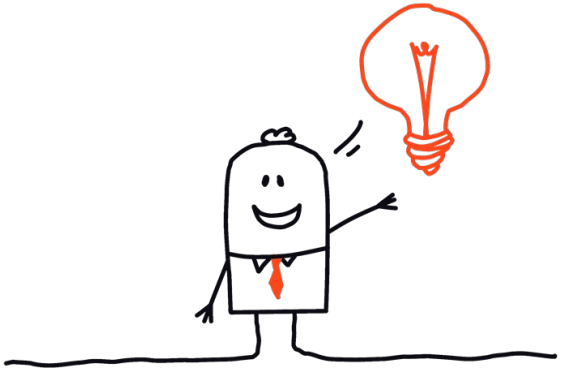
- Pregrado: Ingeniería Informática PUCP
- Maestría en Ciencias de la Computación
  - Binghamton University (SUNY)
  - Becada de Fulbright
  - Grace Hopper Celebration
  - LxMLS
- Estudiante de doctorado



Mi experiencia

- Experiencia académica:
  - IAPUCP
  - Multimedia Research Lab
  - Machine Learning Research Group
- Experiencia profesional:
  - Belcorp, Ericsson
  - Socosoft, BCP
  - RIMAC (jun-agosto 2019)

Agenda

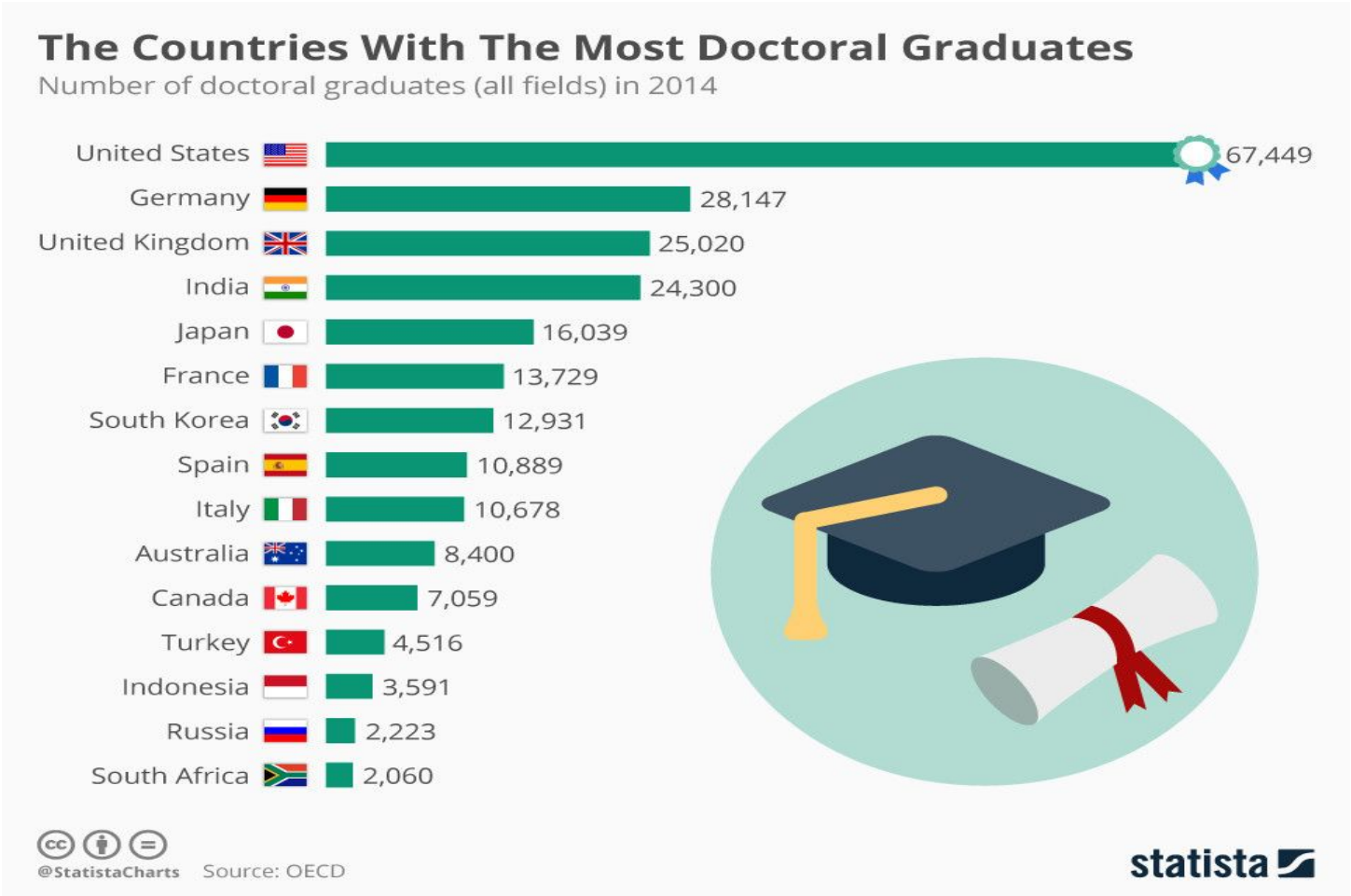


- PhD yo?
  - Innovación y Desarrollo
- Deep Learning y los datos no estructurados
  - Diferencias entre IA, ML y DL
  - Lenguaje Natural (NLP)
  - Visión Computacional (Computer Vision)
  - Señales
- Explosión de Machine Learning (slides de Ian Goodfellow - AAAI19)
- Ética y bias en IA
  - Interpretabilidad y explicatividad

PhD yo?

- ¿Cuántos doctores hay en Perú?
  - Perú: ~4,000
  - Chile, Argentina, España: ~ 17,000 a 50,000
- ¿Cuántos doctores en Computer Science?
- ¿Se necesita un doctorado para hacer investigación/innovación?
- LLamados especialmente a la investigación y docencia.
- ¿Dónde? Industria, Academia

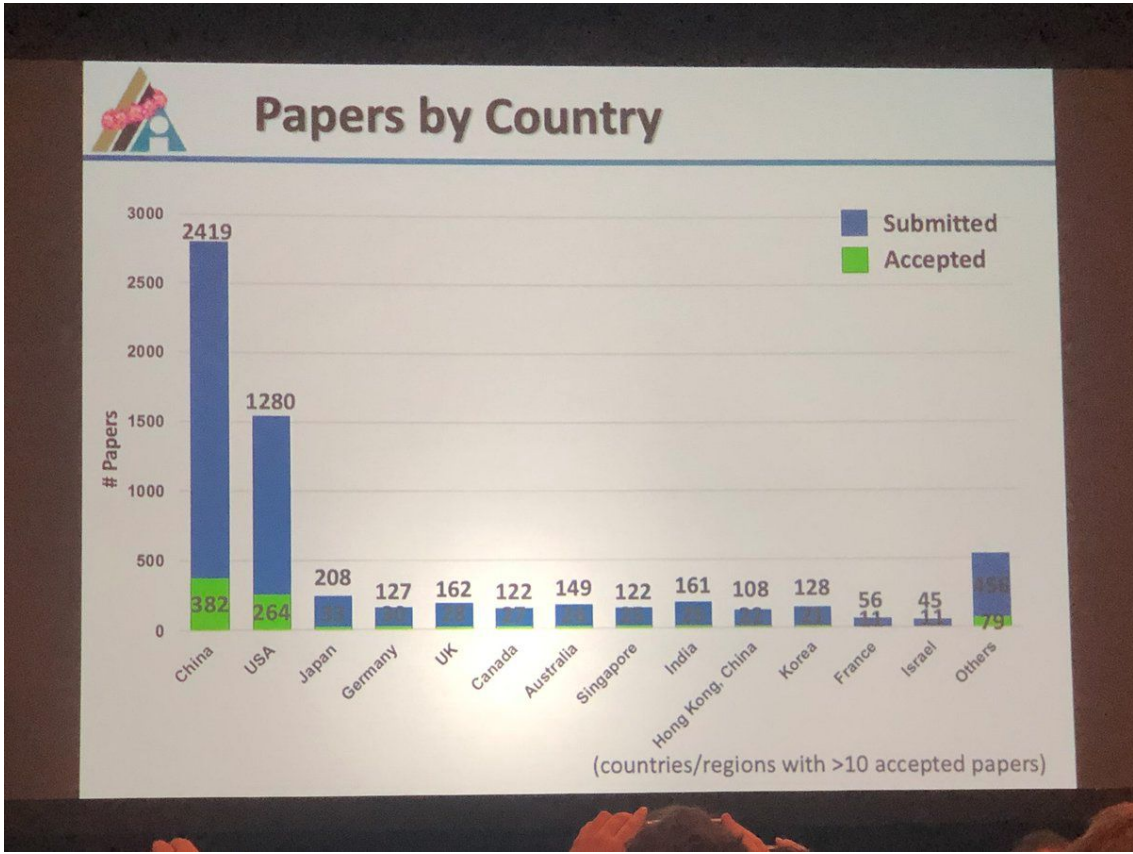
https://www.elmundo.es/f5/campus/2017/10/25/59ef61f146163f721b8b465d.html



https://www.weforum.org/agenda/2017/02/countries-with-most-doctoral-graduates/

Conferencias en IA

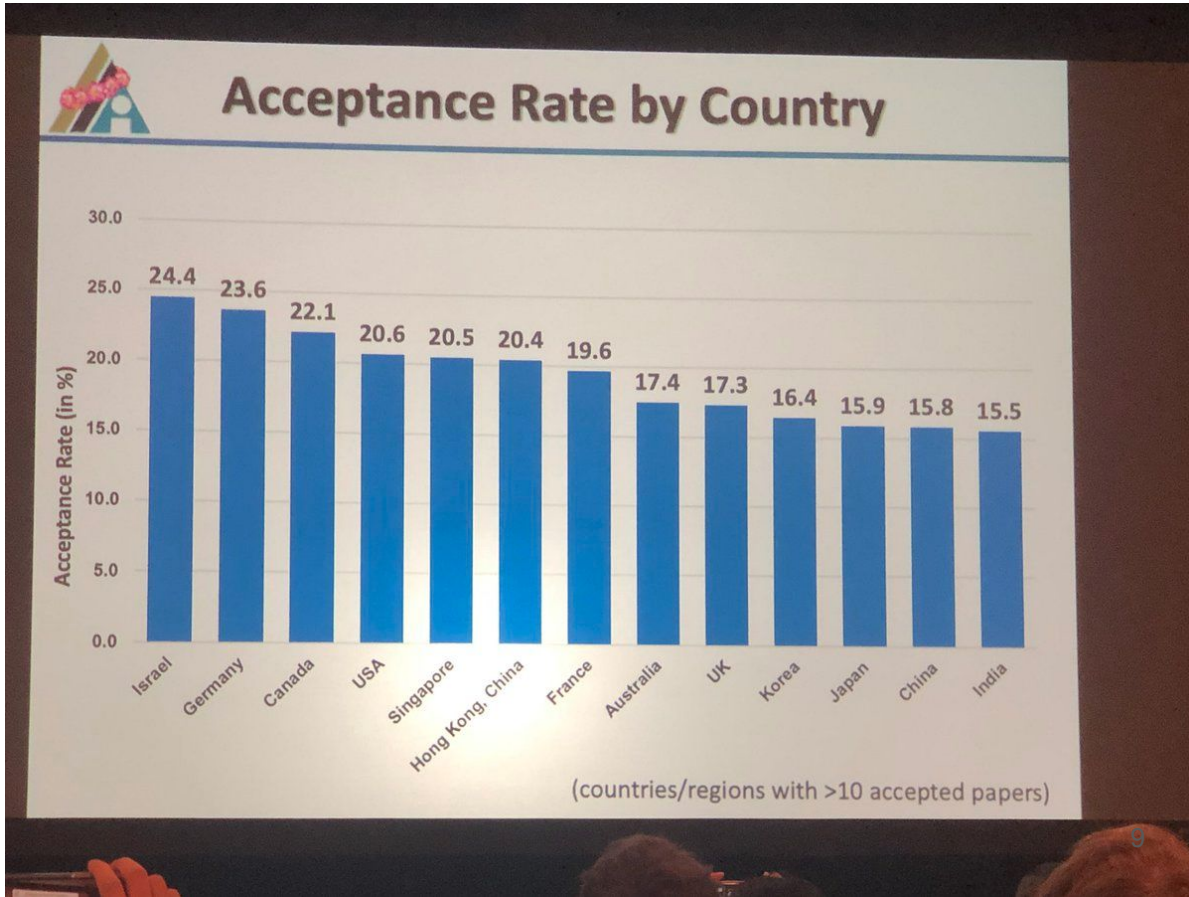
- ¿Qué % de papers proviene de Latinoamérica?
- Top: China, USA, Japón Alemania, Reino Unido Canadá, … , India, Israel



Source: The Thirty-Third *AAAI* Conference on Artificial Intelligence (AAAI 2019) - [@DrXingyuZhao]

Conferencias en IA

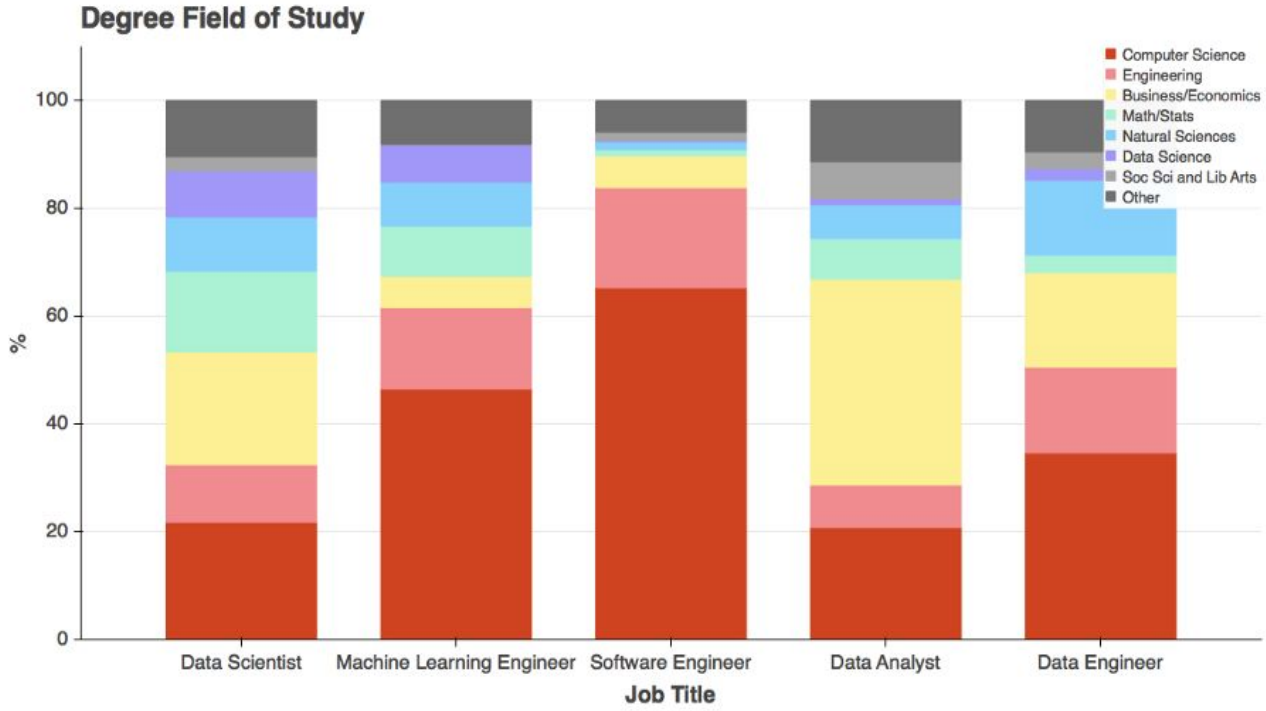
- Ratio de aceptación por país:
  - Israel
  - Alemania
  - Canadá
  - USA
  - …
  - China
  - India



Source: The Thirty-Third *AAAI* Conference on Artificial Intelligence (AAAI 2019) - [@DrXingyuZhao]

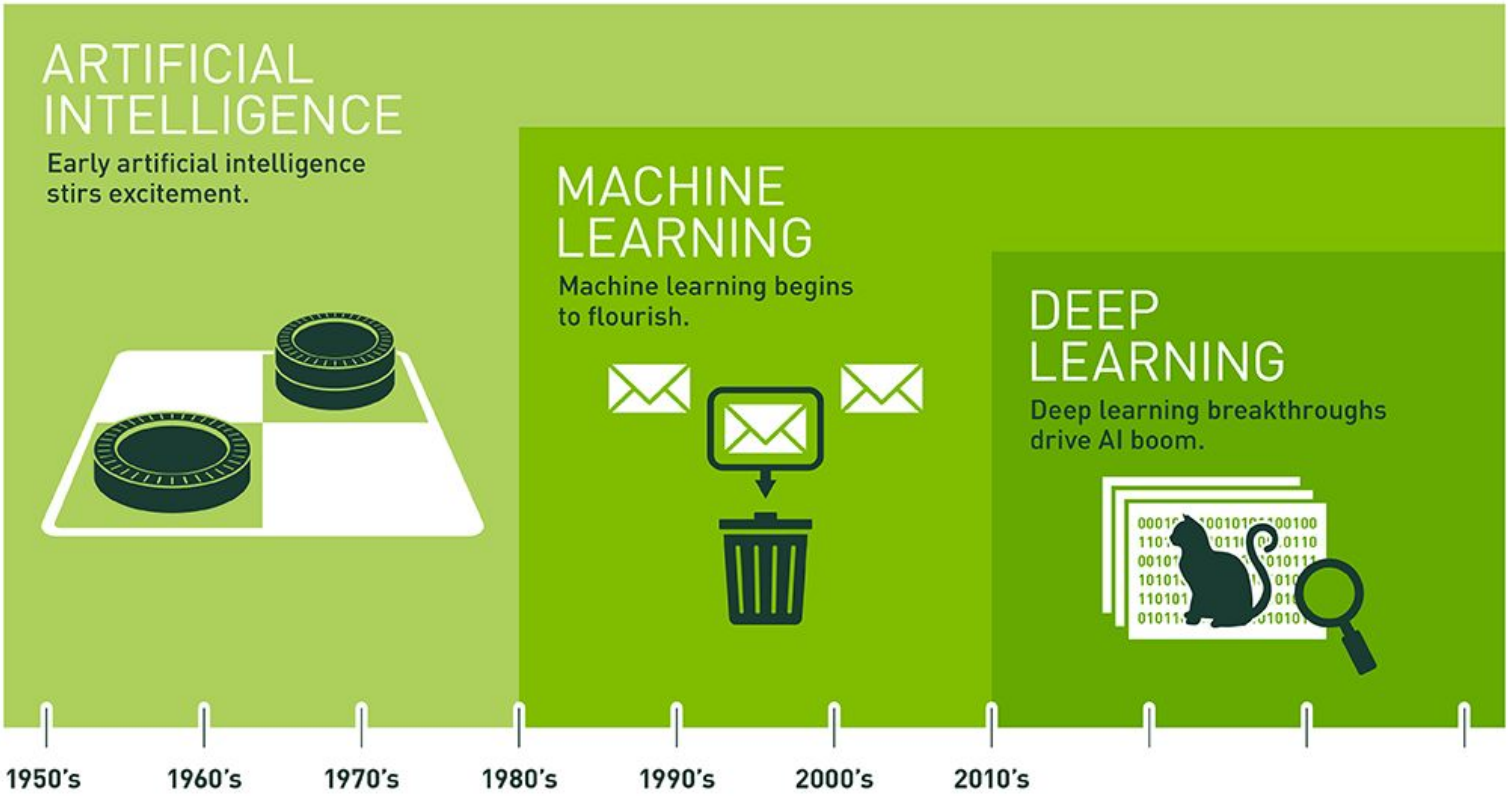
¿De dónde vienen los Data Scientist?

- Datos estructurados
  - Prospección
  - Financieros
- Datos No estructurados
  - Texto
  - Sonido
  - Imágenes
  - Señales



Source: https://medium.com/indeed-engineering/where-do-data-scientists-come-from-fc526023ace

Deep Learning y los datos no estructurados



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Source: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

Lenguaje Natural - Tareas

- Traducción a máquina (Machine Translation)
- Question/Answering
- Summarization
- Paraphrasing
- Generación de diálogo

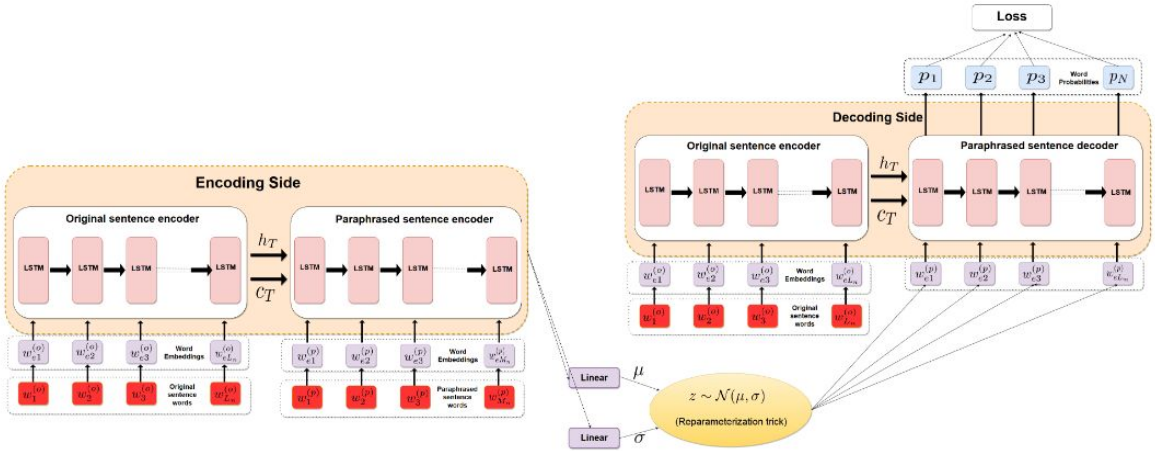
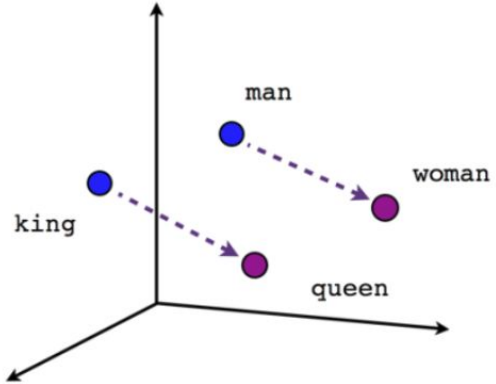


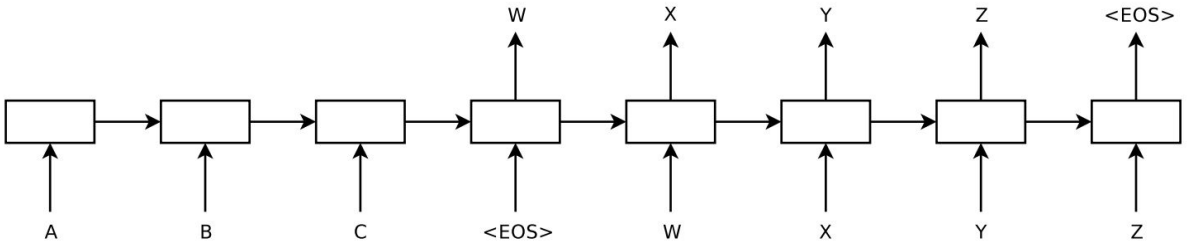
Figure 2: The block diagram of our VAE-LSTM architecture for paraphrase generation

Lenguage Natural - Modelos

- Word Embeddings
  - Glove
  - Fast
  - Elmo
- Generación de texto
  - Tranformers, Bert, GPT2
  - Odelos Sequence2sequence



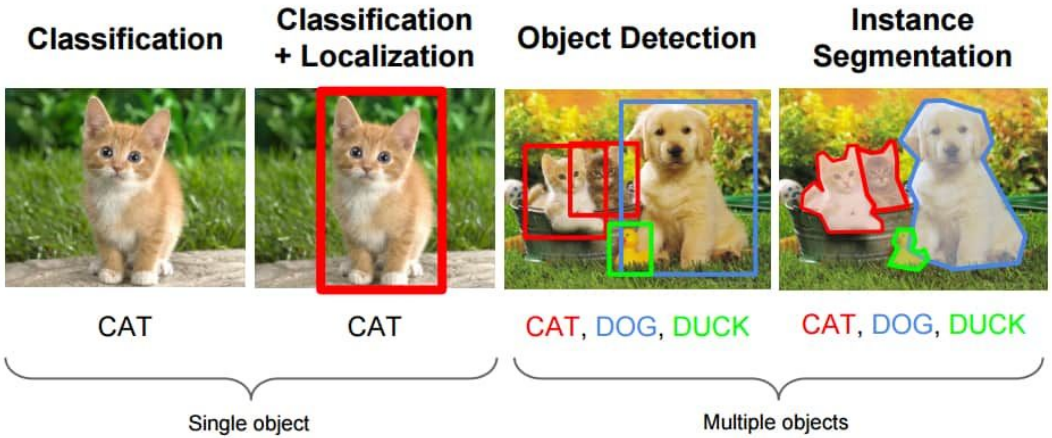
Male-Female



Imágenes

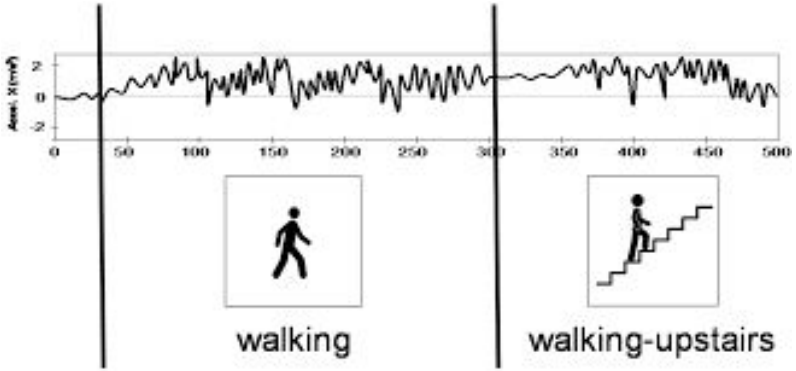
- Reconocimiento de objetos
- Segmentación
- Captioning (leyenda/título)
- Detección de acciones

Source. Ran Xu, Priyanshu Agarwal, Suren Kumar, Venkat N. Krovi, and Jason J. Corso Combining Skeletal Pose with Local Motion for Human Activity Recognition



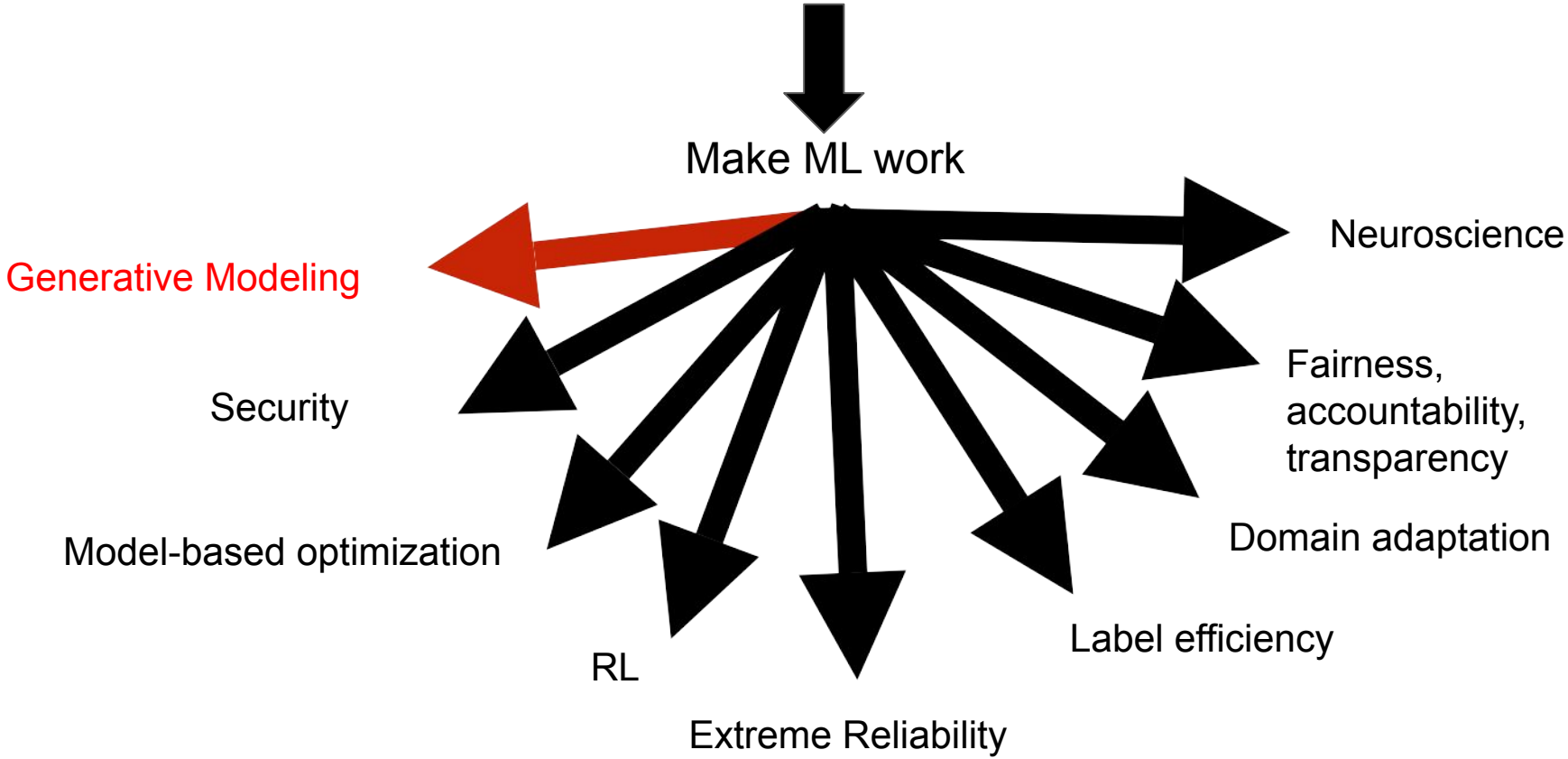
Señales

- Sensores de actividad
- Reconocimiento de voz
- Asistentes virtuales
  - Dominio abierto
  - Dominio cerrado (tareas específicas)



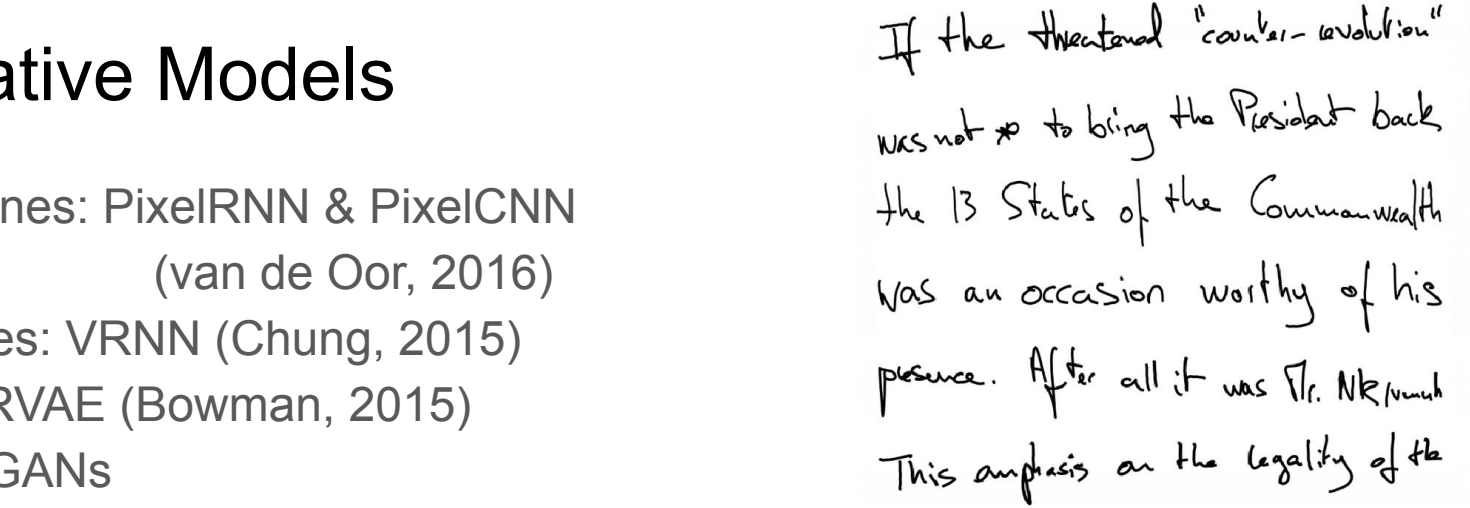
La explosión de Machine Learning

A Cambrian Explosion of Machine Learning Research Topics



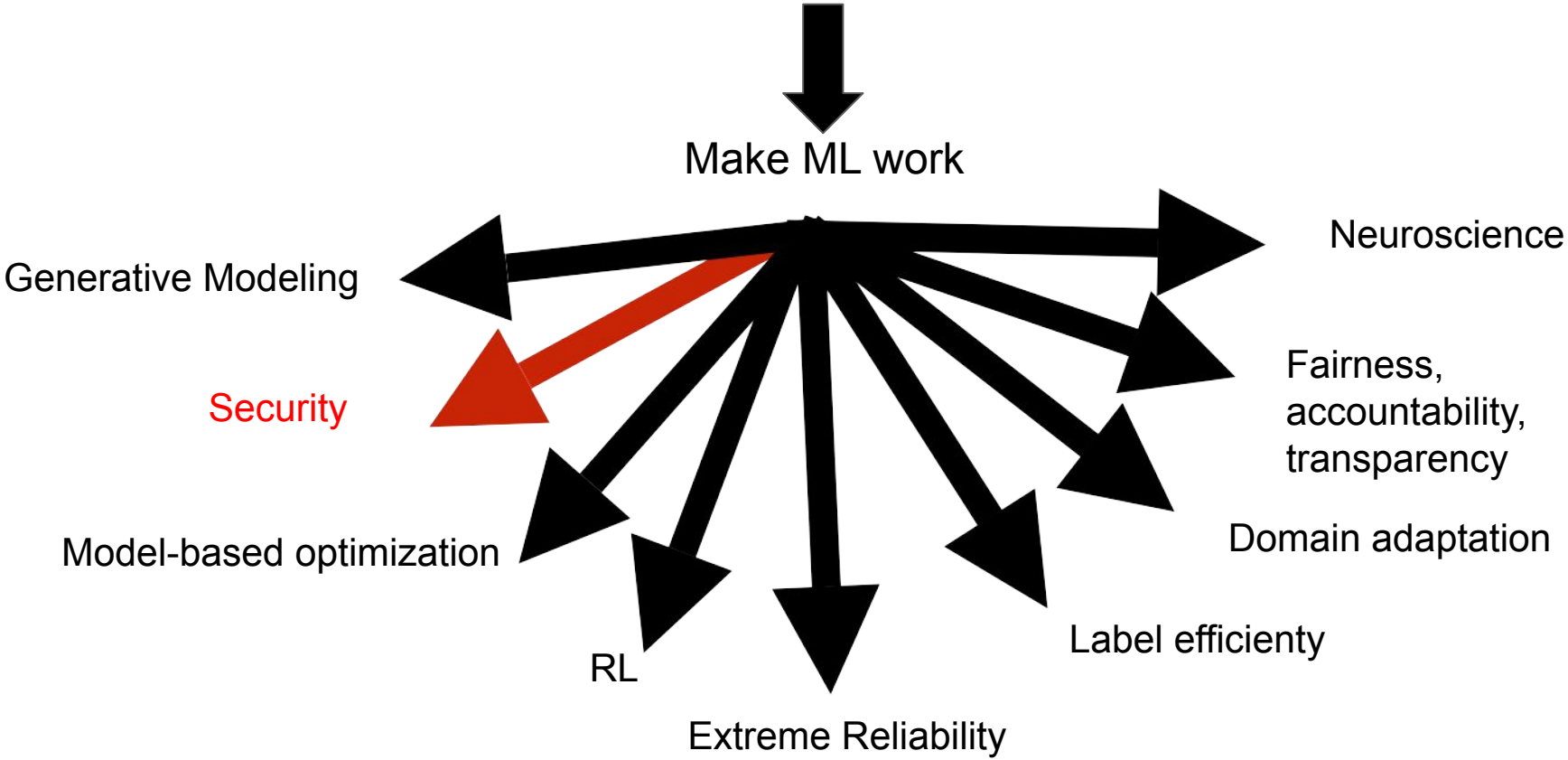
Generative Models

- Imágenes: PixelRNN & PixelCNN
  (van de Oor, 2016)
- Señales: VRNN (Chung, 2015)
- Text: RVAE (Bowman, 2015)
- VAE, GANs



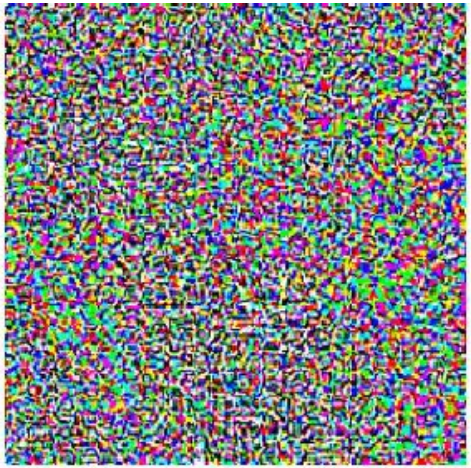
Karras et al, 2017

A Cambrian Explosion of Machine Learning Research Topics



Security: Adversarial examples

$+ .007 \times$

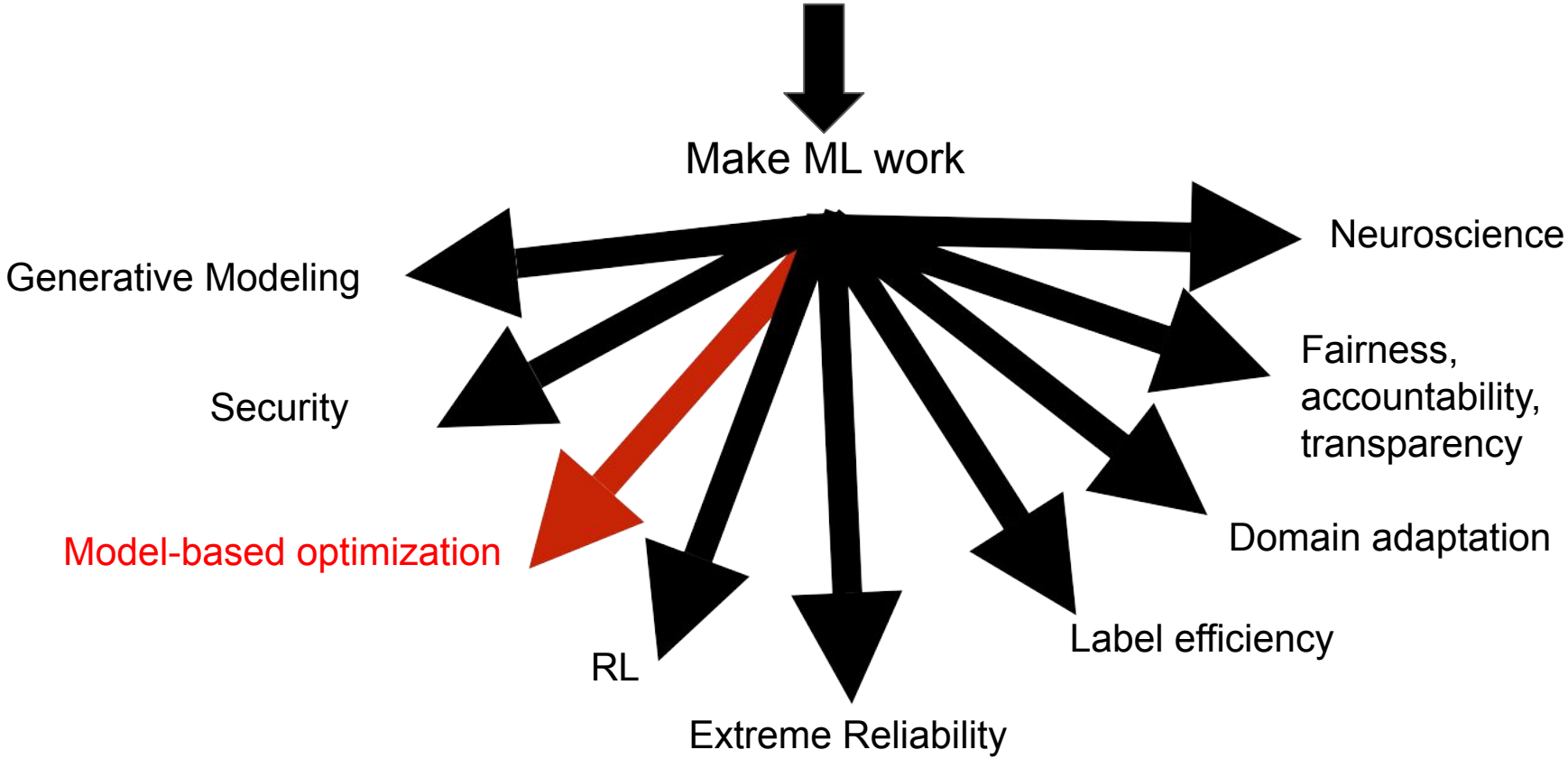


$=$

“Panda” 58%

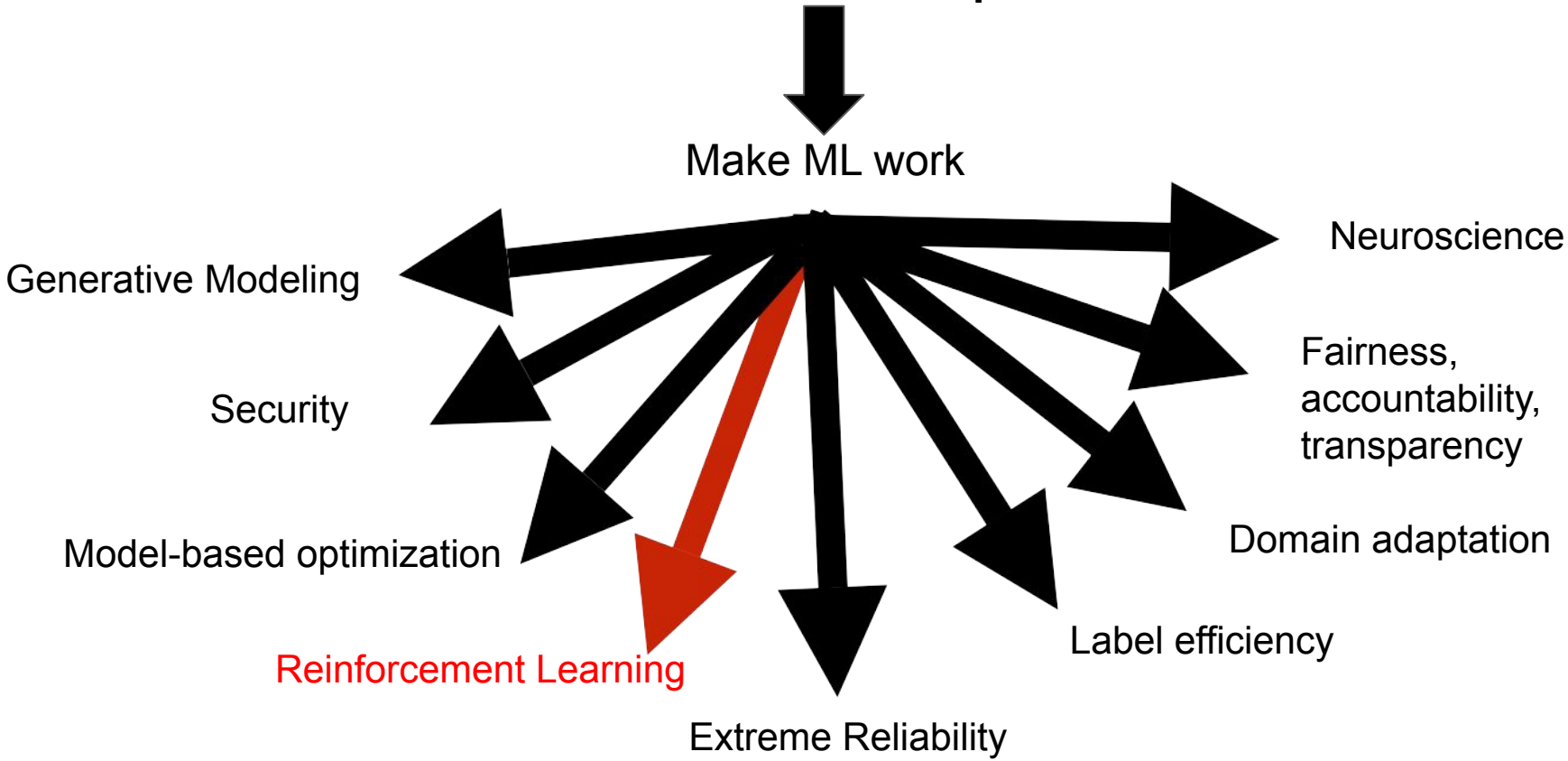
“Gibbon” 99%

A Cambrian Explosion of Machine Learning Research Topics



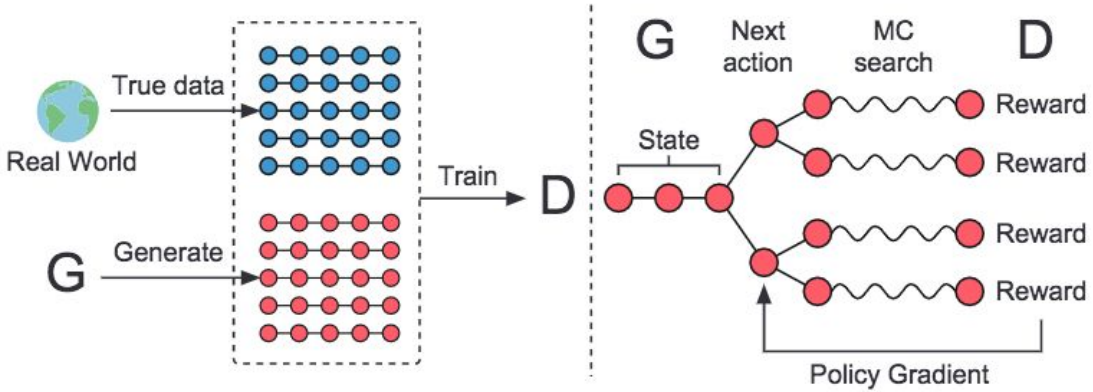
Model-based optimization

A Cambrian Explosion of Machine Learning Research Topics

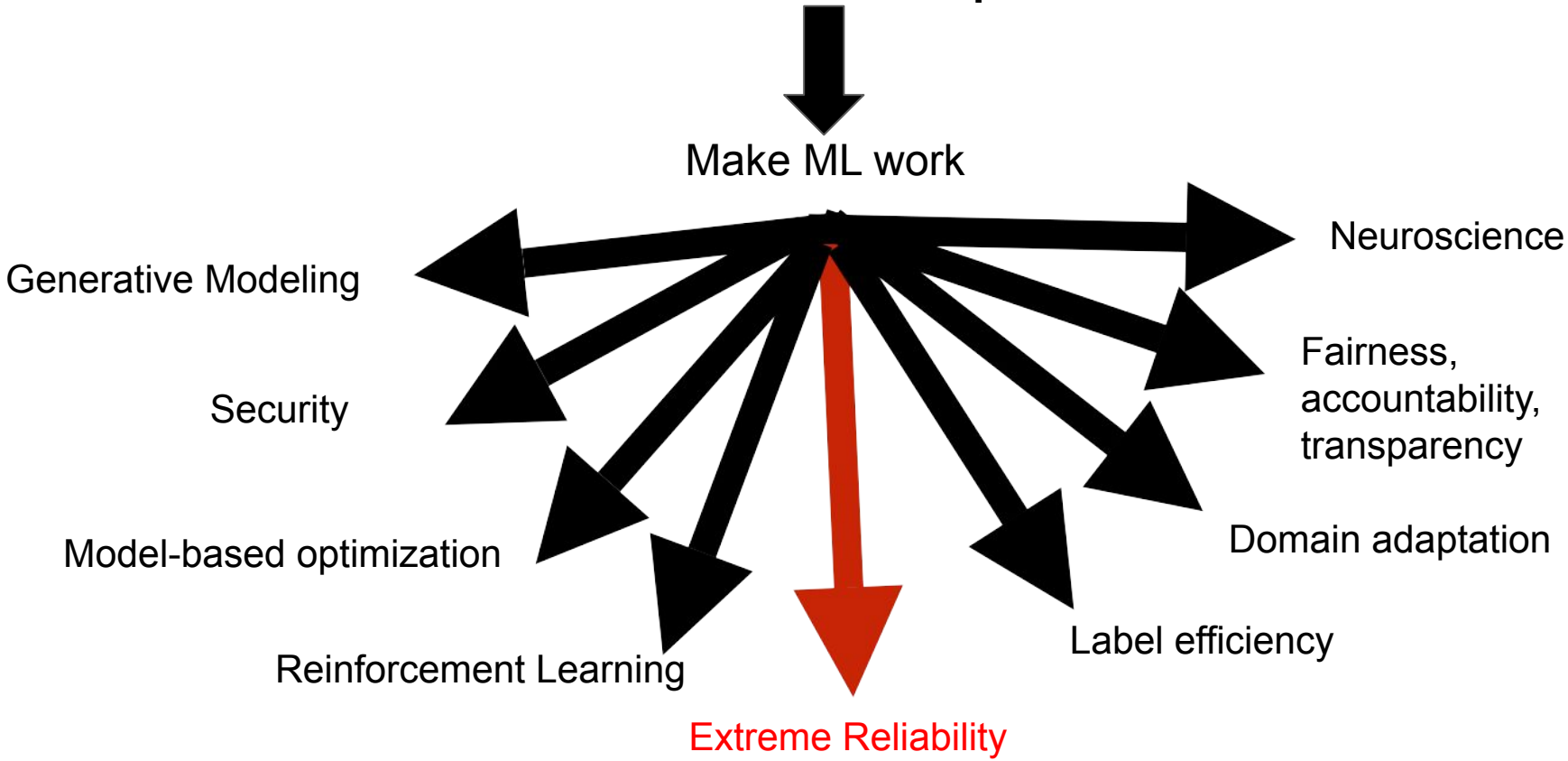


Reinforcement Learning

- Juegos como: Alpha GO, Atari, Starcraft, Dota
- Usado en NLP: Deep Reinforcement Learning for Dialogue Generation (Li, 2017), SeqGAN (Yu et al, 2017), MaskGAN (Fedus, 2018)



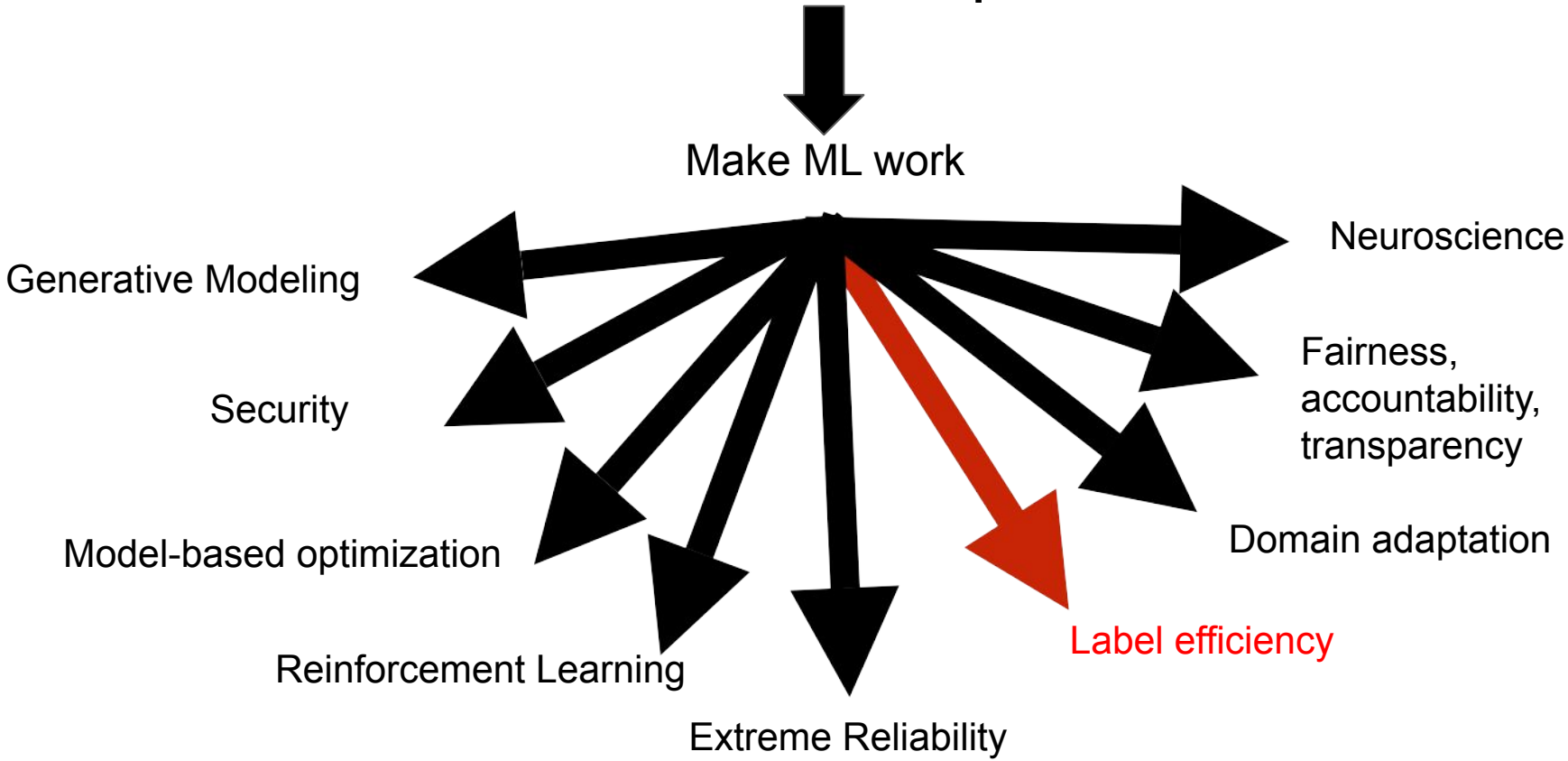
A Cambrian Explosion of Machine Learning Research Topics



Extreme reliability

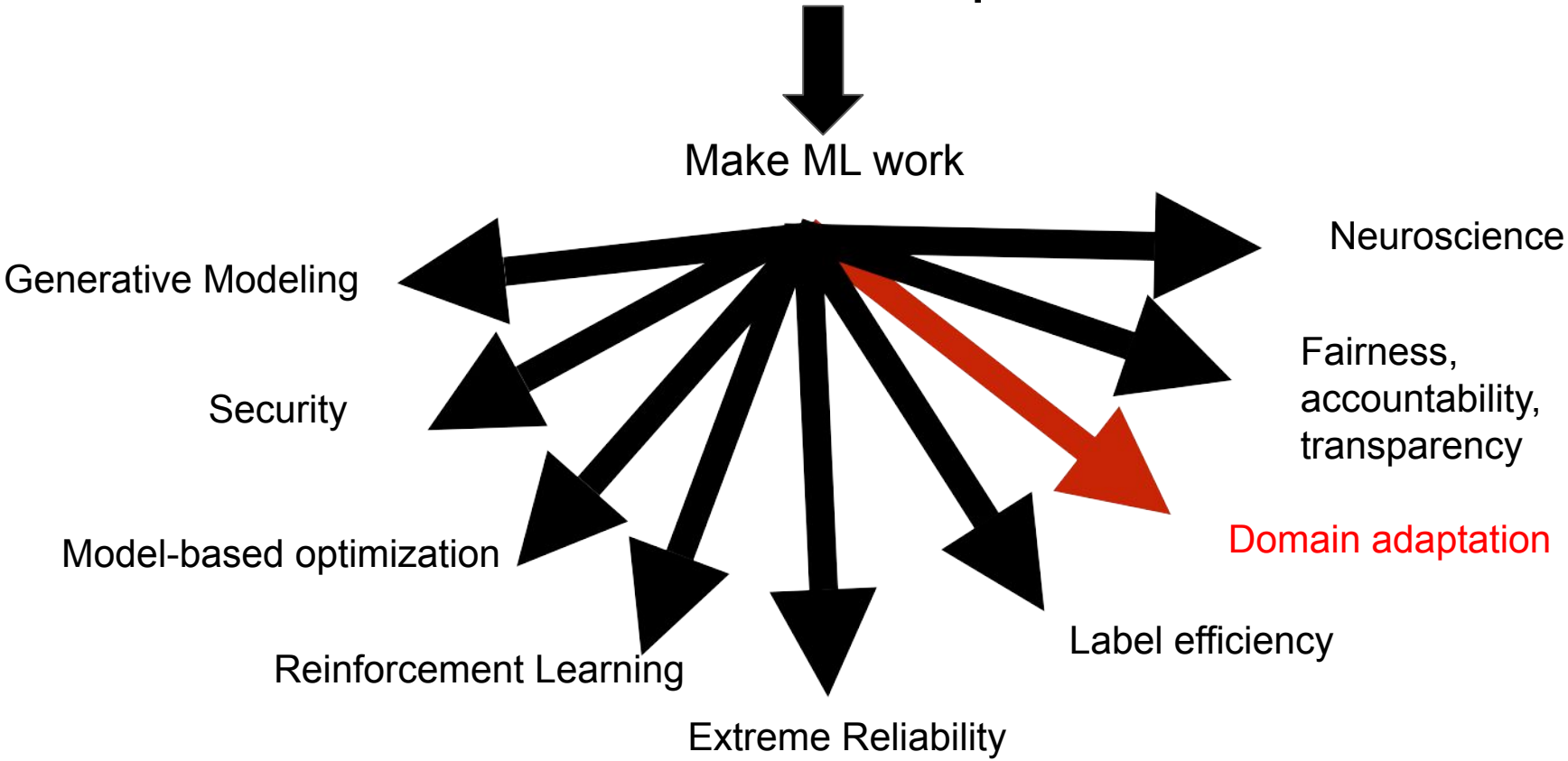
- We want extreme reliability for
  - Autonomous vehicles
  - Air traffic control
  - Surgery robots
  - Medical diagnosis, etc.
- Adversarial machine learning research techniques can help with this

A Cambrian Explosion of Machine Learning Research Topics



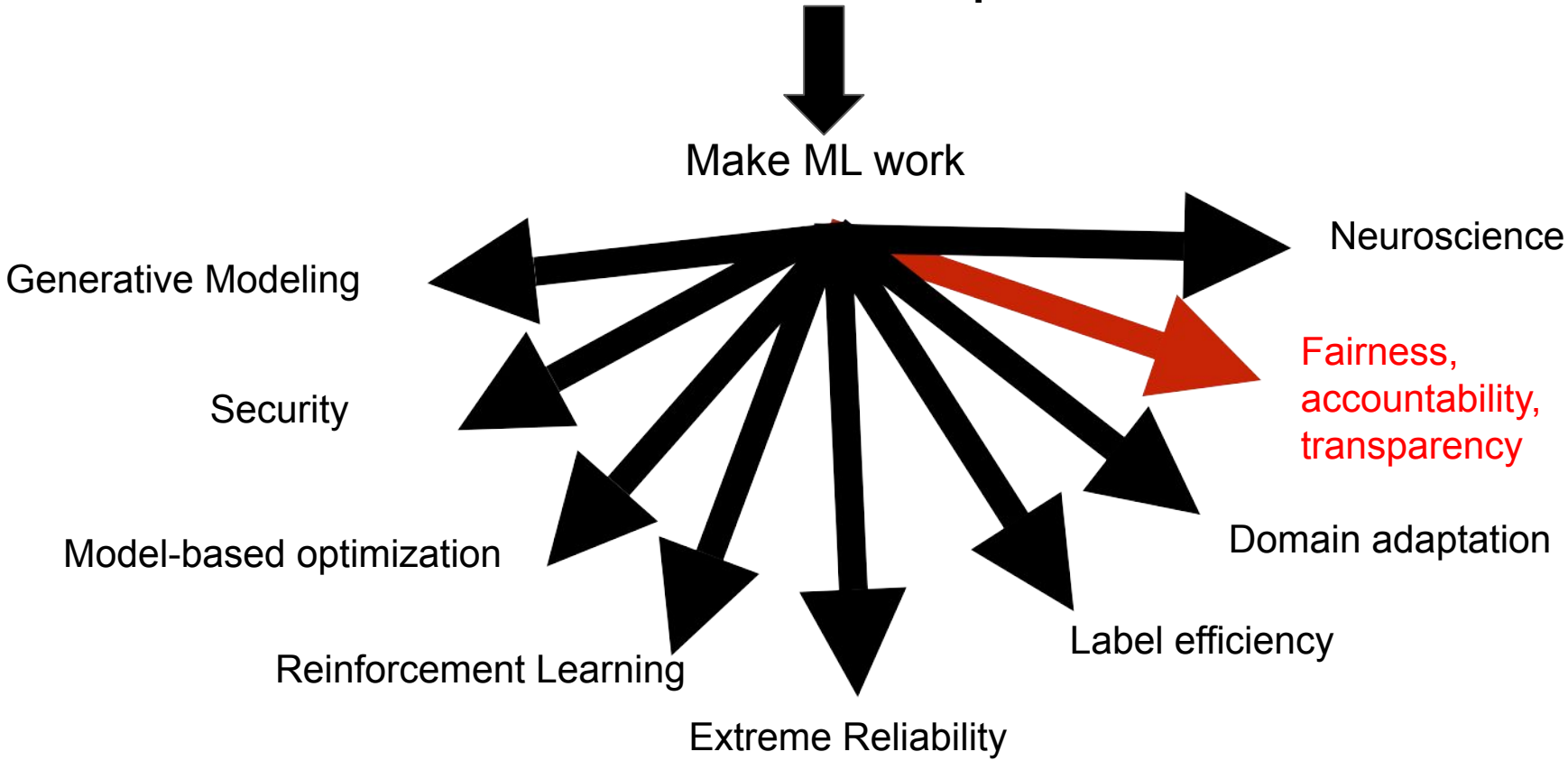
Label Efficiency

A Cambrian Explosion of Machine Learning Research Topics



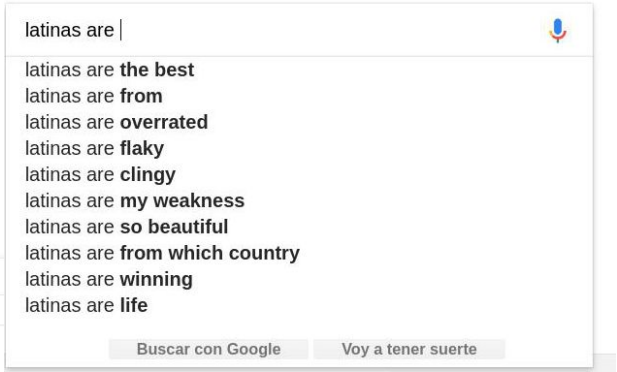
Domain adaptation

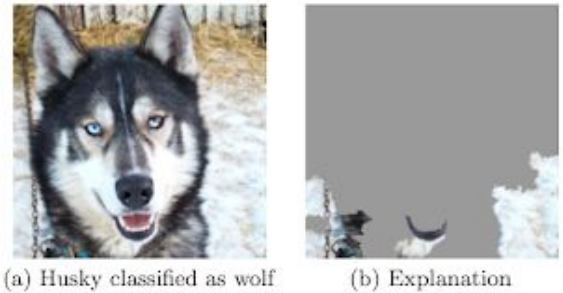
A Cambrian Explosion of Machine Learning Research Topics

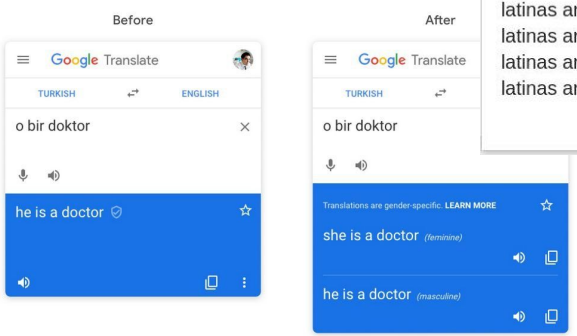


Fairness, accountability and transparency

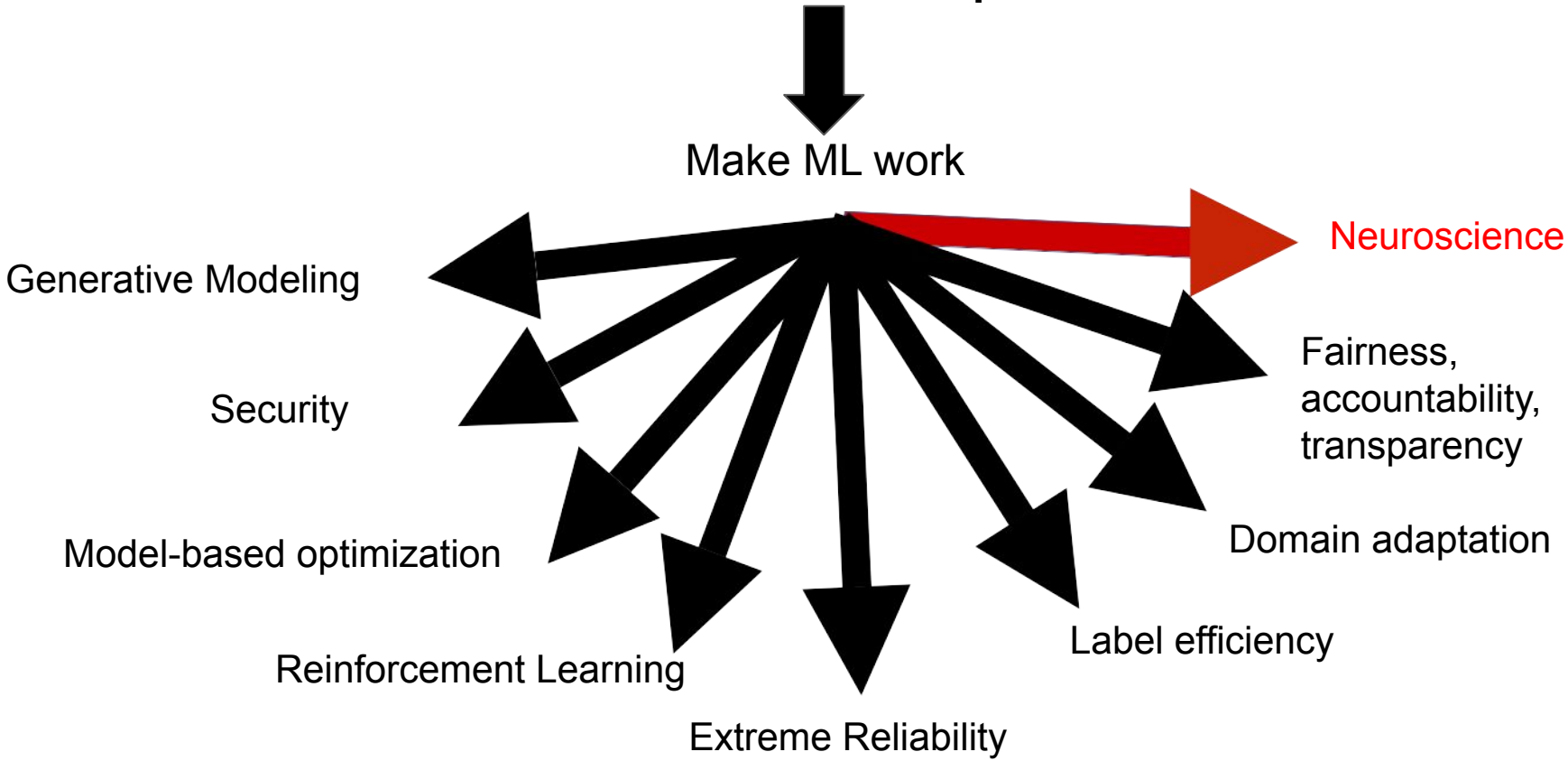
- Buscar en Google:
  - Las mujeres latinas, blancas, negras son …
  - En inglés y español
- Is bias always incorrect?
- Interpretability & explainability
- Traducción de doctor de turco a inglés




(a) Husky classified as wolf (b) Explanation



A Cambrian Explosion of Machine Learning Research Topics



Adversarial examples that affect both computer and time-limited human vision

25%
snake

67%
snake

Elsayed et al 2018

Otros campos a investigar

- Neuroscience
- Quantum Computing
- NLU (natural language understanding)
- NLI (natural language inference)
- Meta-Learning

Miscelánea & Recomendaciones

- Saber buscar papers
- Aprender a replicar papers
- Hinton/Bengio: seguir tu intuición, probar y descartar. Ninguna idea es muy loca
- Documentar (README files, tutorials, blogs)
- Aprender Linux
- “Ensuciarse las manos”

Más fuentes de información

- Algunas personas que seguir en Twitter:
  - Yann LeCun
  - Ilya Sutskever
  - Oriol Vinyals
  - Rachel Thomas
  - Sebastian Ruder
  - Fei Fei Li
  - François Chollet
  - Chelsea Finn
- Podcasts: Practical AI, MIT Artificial Intelligence
- Conferencias: AAAI, ICML, ICLR, NeurIPS, CVPR, ICCV, ACL, NAACL, EMNLP

MUCHAS GRACIAS POR VENIR

WWW.DATASCIENCE.PE