The Challenges of Fundamental Research in Trustworthy AI

ANITI days

Mathieu Serrurier (chaire CALM)

November 24th, 2024

Why Trustworthy AI is Essential?

AI Act: AI have to be

- Safe and transparent
- Ethically aligned
- Respectful of fundamental rights

Critical tasks

- Autonomous systems make critical decisions (e.g., self-driving cars, medical diagnoses).
- Failures or biases can lead to irreversible harm or litigation risks.
- *Certification*

ANITI

10 of ANITI's 19 chairs focus on the fundamentals of AI ...

10 of ANITI's 19 chairs focus on the fundamentals of AI …

… Do we **really** need **so much** fundamental research in AI

?

ANITI

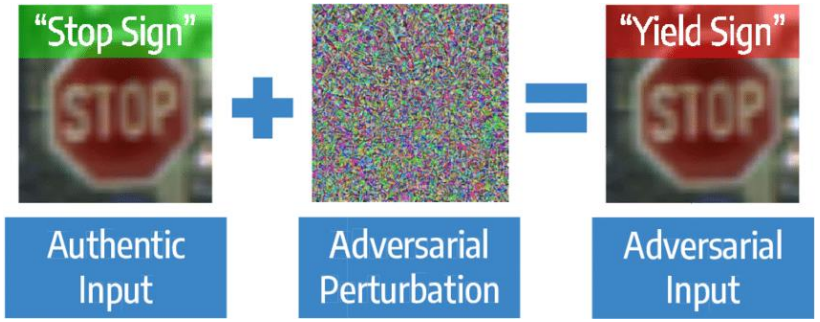
Theoretical Limits of Deep Learning

Deep Learning **works**, but … success often defies traditional assumptions

- **Overparameterized** models outperforming simpler ones.
- **Non-convex** optimization yet effective solutions.

Structural Challenges

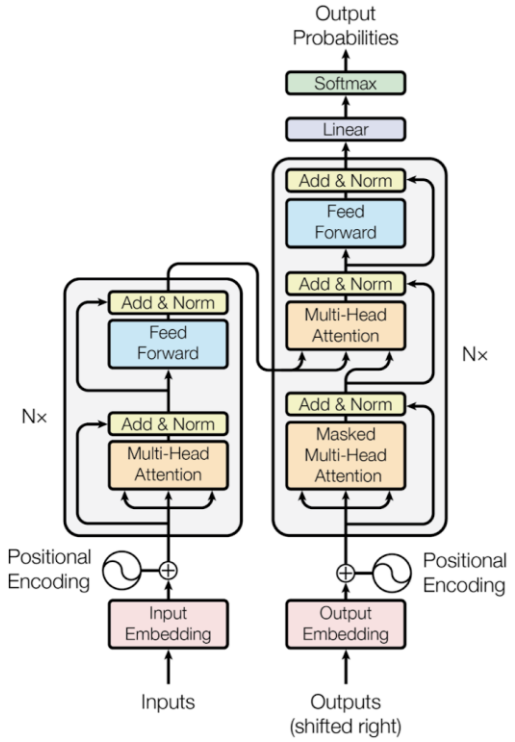
- **Adversarial Vulnerabilities**: Minor input changes lead to major errors.



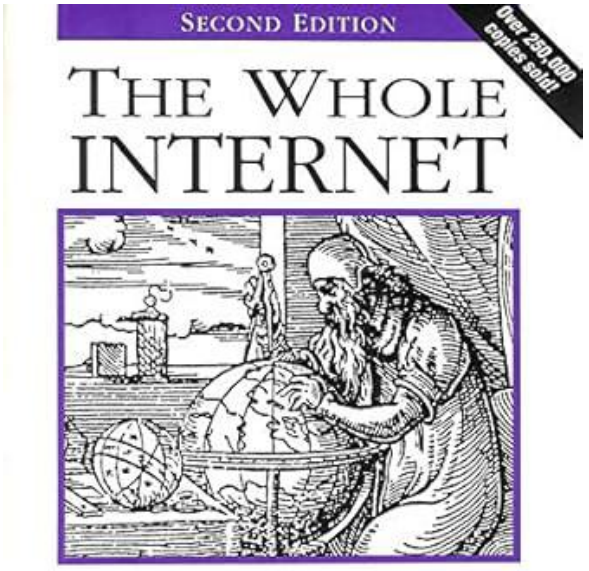
- **Lack of Generalization:** Struggles with out-of-distribution data.
- **Data-Hungry Nature:** Performance hinges on massive datasets.

ANITI

Empirical Success, Theoretical Mysteries



+



=

How can we ensure the safety of an AI-based model decisions?

What guarantees can we expect from fundamental research in AI?

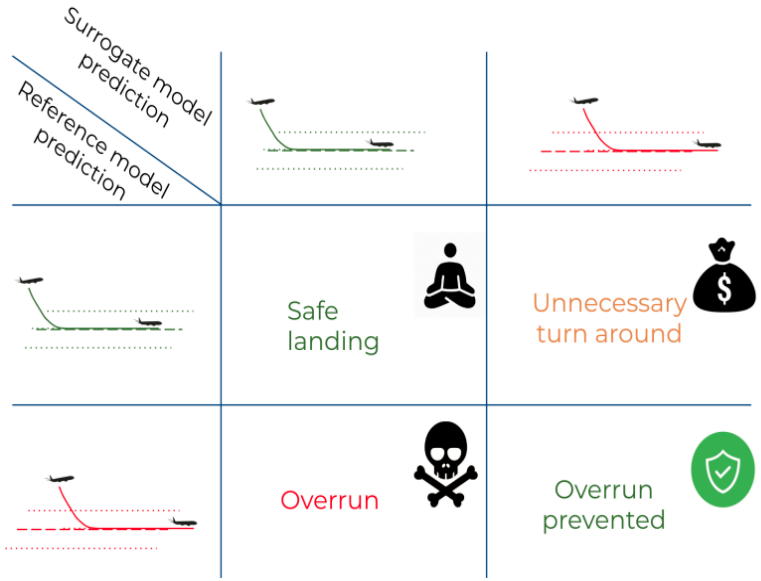
Guarantees on the Data:	Guarantees on the Models:	Guarantees on the Training Algorithm:
• Fairness • Frugality • OOD • …	• Explainability • Robustness • Certified predictions • Uncertainty quantification • Physically informed • …	• Efficiency • Convergence • Soundness • …

ANITI

An industrial usecase: Braking Distance Estimation

23/09/2022: a Boeing overrun in Montpellier

EASA: 41 accidents involving small non commercial airplanes happen during landing (1991-2017)



… Do we **really** need **so much** fundamental research in AI ?

YES

ANITI