AI 101:

An Opinionated Computer Scientist’s View

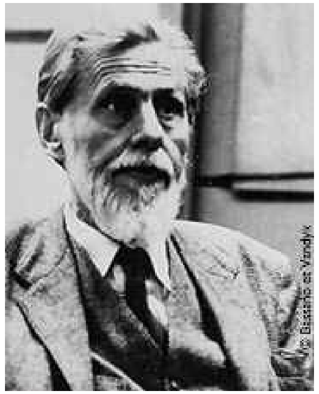
Ed Felten

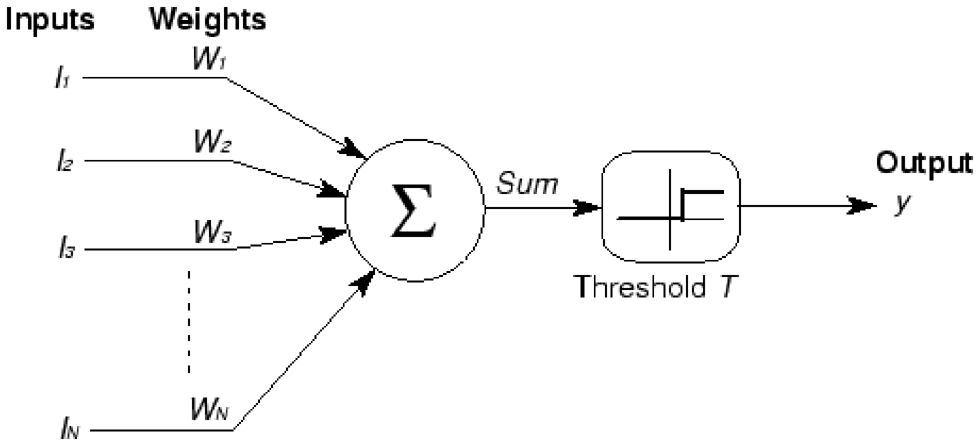
Robert E. Kahn Professor of Computer Science and Public Affairs
Director, Center for Information Technology Policy
Princeton University

A Brief History of AI

History of AI: Birth of the Field

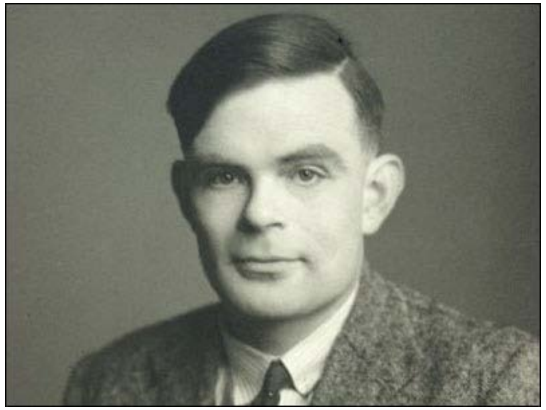
McCulloch-Pitts 1943 :
“A Logical Calculus of the Ideas Immanent in Nervous Activity”





Key idea:
mathematical structures
inspired by the brain
can do complex logical reasoning

History of AI: The Turing Test



Turing 1950: *Computing Machinery and Intelligence*
"I propose to answer the question: can machines think?"

"Imitation Game" or "Turing Test":
Can a machine impersonate a person, in a chat room?

Key ideas:
1. Intelligence is defined by behavior, not internal experience.
2. Goal is to behave as a person would.

History of AI: 1950-2000

Slow but steady progress

Waves of huge optimism and pessimism
(despite more steady progress in underlying technology)

Grand challenges remained
interpreting complex inputs: image and speech recognition
natural language: translation, summarization
complex games: Go, poker
safety-critical control: driving a car

History of AI: Sudden Acceleration (2010-now)

Surge of progress – superhuman performance on grand challenges

Driven by combination of
big datasets
better algorithms
bigger, faster computers

Big tech companies investing heavily in AI

Popular interest in AI

Three Lessons Learned About AI

AI Lesson #1

AI is not a single thing – it’s different solutions for different tasks.

Narrow AI

Focus on a specific narrow task
Develop task-specific solutions

Steady progress
Growing excitement

General AI

Usable for any cognitive task
General, adaptive intelligence
What humans have

Not much progress
Excitement (+ some hysteria)

Narrow AI is useful for if you want to make money.
General AI is useful if you want to make movies.
-- paraphrasing Dave Honey

AI Lesson #1

AI is not a single thing – it’s different solutions for different tasks.

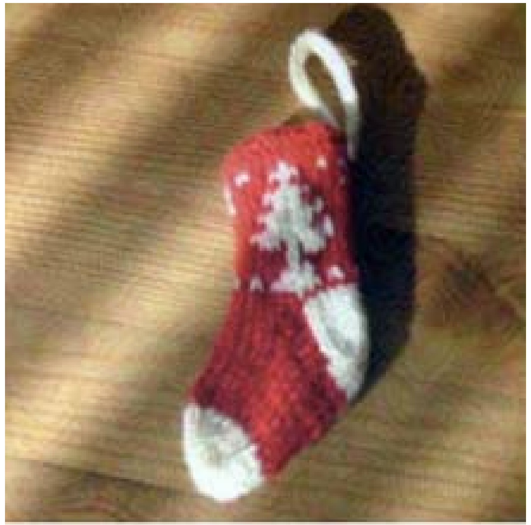
AI will surpass us at different times for different tasks.

It might be hard to predict when a particular task will be automatable.

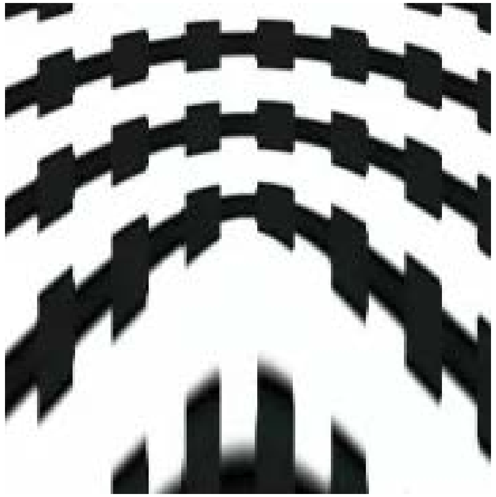
AI Lesson #2

Successful AI doesn’t think like a human — it’s an *alien* intelligence.

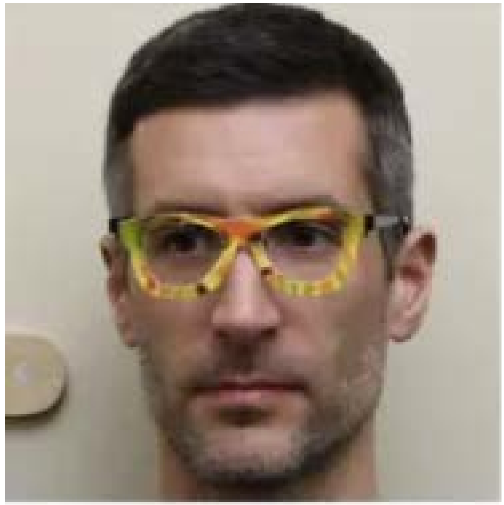
Some AI Errors



Indian elephant

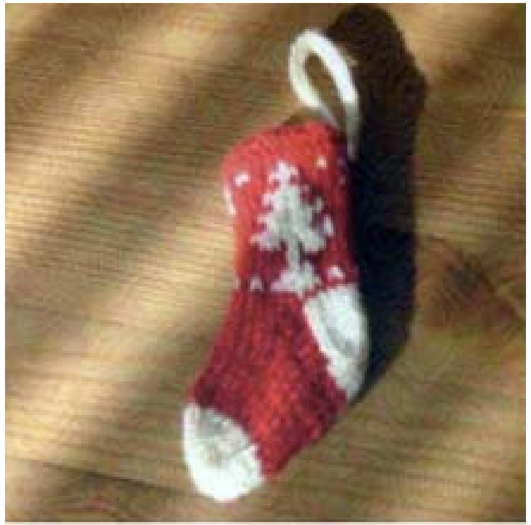


Assault rifle

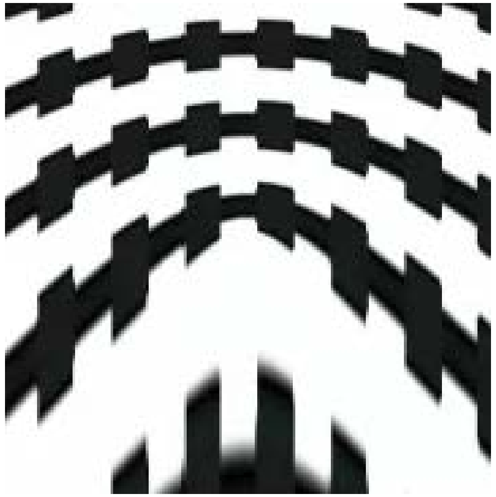


Milla Jovovich

Some AI Errors



Indian elephant



Assault rifle

AI Lesson #2

Successful AI doesn’t think like a human — it’s an *alien* intelligence.

AI’s errors won’t be like human errors.

Advanced AI will have a different “style” than humans.

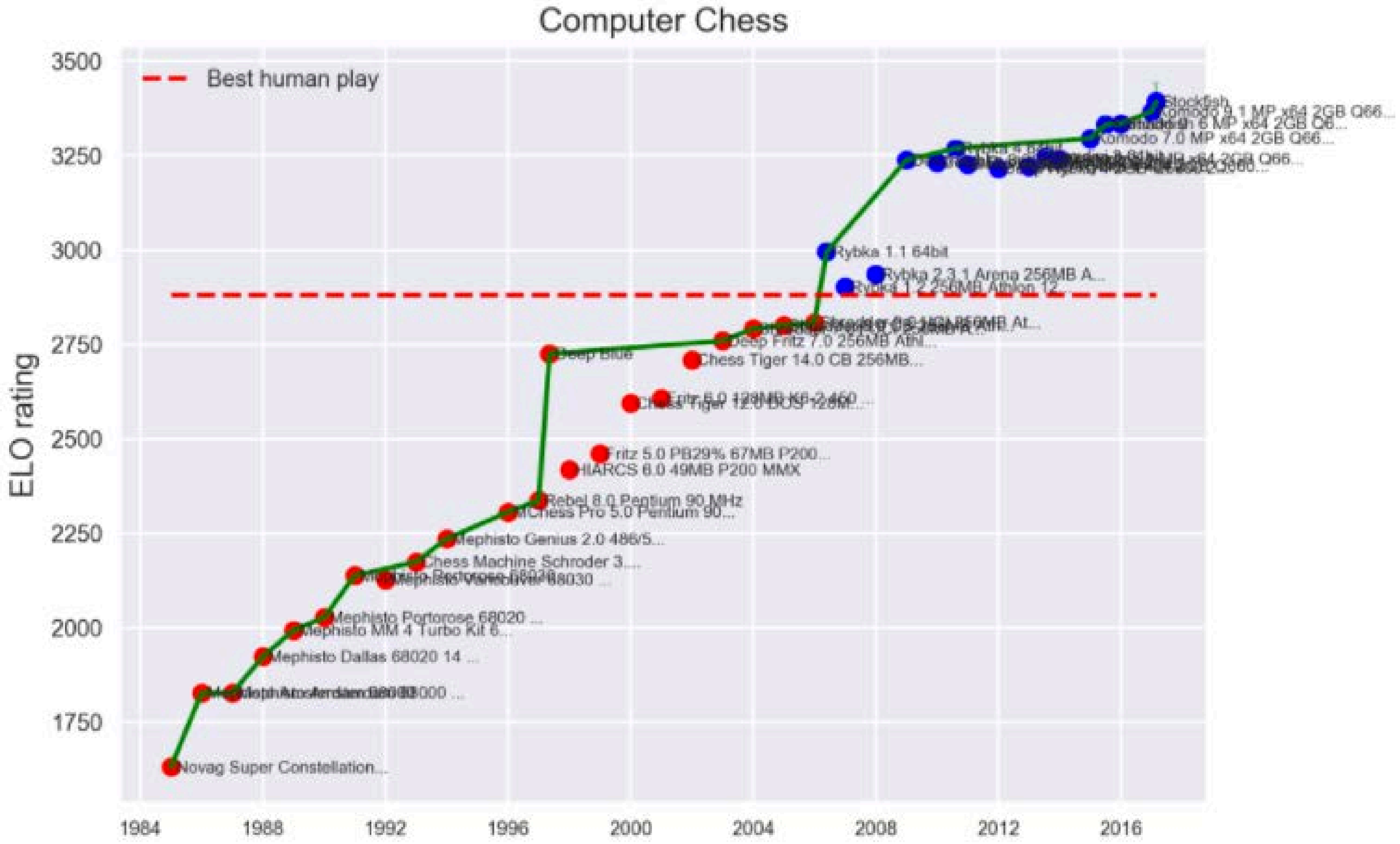
What is easy for AI might be difficult for humans, and vice versa.

Effective machine-human teaming may be valuable—but hard to get right.

AI Lesson #3

On many cognitive tasks, more engineering effort or more data translates into better AI performance.

Steady Progress by Effort (Chess)



AI Lesson #3

<u>On many cognitive tasks, more engineering effort or more data translates into better AI performance.</u>

Machines are worse than humans at learning from experience,
but a machine with lots of data has much more experience to learn from.

On Explainability

People often say that AI results aren’t explainable.

What does this mean?

AI systems are much more transparent, in detail, than a human brain.

Four Flavors of Explainability Complaints

Non-transparency: Explanatory information exists but is being withheld.

Complexity: Detailed explanatory information exists, but nobody can find a simple holistic summary of the algorithm’s behavior.

Non-intuitiveness: The system discovered a statistically valid rule, and we understand that rule, but we don’t know why the rule is effective.

Lack of justification: We understand how the system works, but we want a justification for why the outcomes are fair or reasonable.

AI 101:

An Opinionated Computer Scientist’s View

Ed Felten

Robert E. Kahn Professor of Computer Science and Public Affairs
Director, Center for Information Technology Policy
Princeton University