

Artificial Intelligence and the Brave New World in Finance

Markus K. Brunnermeier

G20 Global Financial Stability Conference

27th of October, 2026

AI: Revolution or Simply Another Innovation?

- **Scientists:** AI may fundamentally disrupt mankind
 - *"These things do **understand**. And because they understand, we need to think hard about what's going to happen next. And we just don't know."*
Geoffrey Hinton, Nobel Laureate in Physics, 2025
- **Economists:** Fundamentally disrupts
 - Job displacements
 - Growth long-term
- **Finance and money:** yet research treats AI as a standard financial innovation?
 - Finance will be one of the **first fields disrupted** since it is data and information driven
 - *"Within a decade, no human on the planet will **understand** the financial system."*
Yuval Noah Harari, The Economist, Aug. 2026

The Big AI Challenge: From Asymmetric Info to Understanding

- **Bankers build trust** ... and so do Central Bankers
 - Delegation to institutions (run by humans) to **reduce** commitment and **asymmetric information** frictions
 - Institutional trust arrangement $\Rightarrow$ it is enough to trust someone who understands
- $\Rightarrow$ “societal understanding” via trust
-

The Big AI Challenge: From Asymmetric Info to Understanding

- **Bankers build trust** ... and so do Central Bankers
 - Delegation to institutions (run by humans) to **reduce** commitment and **asymmetric information** frictions
 - Institutional trust arrangement $\Rightarrow$ it is enough to trust someone who understands

$\Rightarrow$ “societal understanding” via trust

- AI is revolutionary technology
 - (i) **disrupts** current institutions & **trust** (arrangement)
 - (ii) AI agents have **different/deeper representation** of the world

$\Rightarrow$ instead of an **asymmetric information** an **Asymmetric UNDERSTANDING** problem

What is Understanding?

- **Understanding** $\approx$ (dynamic) representation of the world, language
 - Invariant to deeper microfoundation of representation / does not improve
 - Aspirin, ... (invariant to [Lucas critique](#))

What is Understanding?

- **Understanding** $\approx$ (dynamic) representation of the world, language
 - Invariant to deeper microfoundation of representation / does not improve
 - Aspirin, ... (invariant to **Lucas critique**)
- **Shared understanding** $\approx$ shared language
 - $\Rightarrow$ allows each party to model other parties' reasoning and motives,
residual ignorance is due to information
addressable by **communication, monitoring, contracting**, ... financial institutions

What is Understanding?

- **Understanding** $\approx$ (dynamic) representation of the world, language
 - Invariant to deeper microfoundation of representation / does not improve
 - Aspirin, ... (invariant to **Lucas critique**)
- **Shared understanding** $\approx$ shared language
 - $\Rightarrow$ allows each party to model other parties' reasoning and motives,
residual ignorance is due to information
addressable by **communication, monitoring, contracting**, ... financial institutions
- **Societal understanding via trust**
 - Institutional trust arrangement $\Rightarrow$ it is enough to trust someone who understands

Trust

- **Delegation** to institutions (run by humans) to **reduce** commitment and **asymmetric information** frictions
 - Bankers build trust ... and so do Central Bankers
- **Societal** understanding via **trust**
 - Institutional trust arrangement $\Rightarrow$ it is enough to trust someone who understands
- Few citizens understand monetary policy; they trust an institution whose staff does
 $\Rightarrow$ societal understanding

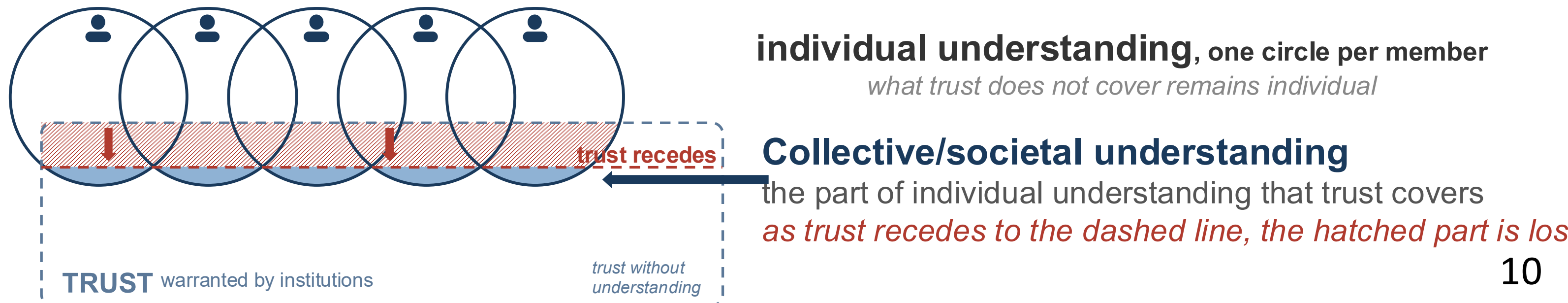
What is Understanding?

- **Understanding** $\approx$ (dynamic) representation of the world, language
 - Invariant to deeper microfoundation of representation (invariant to **Lucas critique**)
- **Shared understanding** $\approx$ shared language
 - $\Rightarrow$ allows each party to model other parties' reasoning and motives,
residual ignorance is due to information
addressable by **communication, monitoring, contracting**, ... financial institutions
- **Societal understanding via trust**
 - Institutional trust arrangement $\Rightarrow$ it is enough to trust someone who understands



What is Understanding?

- **Understanding** $\approx$ (dynamic) representation of the world, language
 - Invariant to deeper microfoundation of representation (invariant to **Lucas critique**)
- **Shared understanding** $\approx$ shared language
 - $\Rightarrow$ allows each party to model other parties' reasoning and motives,
residual ignorance is due to information
addressable by **communication, monitoring, contracting**, ... financial institutions
- **Societal understanding via trust**
 - Institutional trust arrangement $\Rightarrow$ it is enough to trust someone who understands



What is Asymmetric Understanding?

- **Asymmetric:** AI “understands” humans’ understanding, but not vice versa

1. **Non-explainable** can’t retrace process of AI agents’ decisions (black box)
non ← verifiable weaponize → deceive
2. **Misalignment** can’t correctly specify reward function (Goodhart’s Law)
(undetectable, illegibility can be strategic (deceptive))

⇒ instead of an asymmetric information an Asymmetric UNDERSTANDING problem

Roadmap

1. Intro: Trust
2. From Asymmetric Information to **Asymmetric Understanding**
3. Financial Markets with AI Agents
 - Segmentation, ...
4. Central Banking in the Age of AI
 - Monetary Policy as a Lopsided Game btw Central Bank and Market Participants
 - Liquidity Policy (LLoR) and AI Moral Hazard
 - Counterintuitive Shifts in Central Banking Tradeoffs
5. Transition, Resilience,
AI Dependency and Power Concentration $\Rightarrow$ systemic risk
6. Conclusion

3. AI Financial Markets: Instability and Manipulation

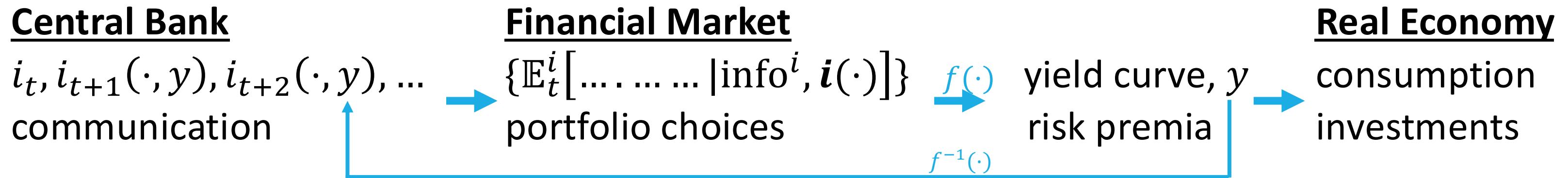
- Warning from succession of flash crashes (due to algo trading)
 - Asymmetric understanding introduced by AI traders makes the **market's reaction function** even **more erratic**
 - Price signal is difficult to “understand”: $f(\cdot): \{\text{info}^i\} \xrightarrow{f(\cdot)} P$ is difficult to invert (“back out”)
 1. Less stabilizing influence of arbitrage (like noise trader risk)
 2. Non-explainable collusion among AI traders (pump and dump, ...)
 3. Rogue trader as a precursor of AI agents:
 - Narrow interpretation of mandate, less sensitivity to risk
- ⇒ Segmentation in markets and circuit breakers?
- **AI worsens the trade off between efficiency and stability**

4. Central Banking: MoPo and Liquidity Policy

- AI changes the strategic interaction inside financial markets and
btw market and central bank;
it creates an understanding asymmetry btw central bank and private agents
- *Current:* CB has a knowledge/observational advantage on state of the economy
market participants have dispersed info^i
- *Future:* AI has an **understanding advantage**:
 - They understand the system, composed of themselves and the CB, better than CB does;
 - AI agent are better in inferring the CB's true objective than the reverse
 - AI market participants understand CB reaction function (best response)
 - CB does **not** understand market's reaction function (has to play max-min strategy)

4a. Monetary Policy with AI Market Participants

- MoPo transmission indirect via help of (informed) financial market as a **game** btw central bank and financial market participants (in a game-theoretic sense)



- Central Bank chooses $i(\cdot)$ -interest rate rule to “**use**” markets
- Market participants forms expectations, try to “**outsmart**” CB (e.g. Soros), and choose their reaction function (portfolio choice) ...
- Hall-of-mirrors problem $i(\cdot, y)$, Stein-Sunderam on gradualism/whispering, ...
- Adversarial: CB commits its balance sheet to *defend* future rates, YCC, caps, spread
Financial dominance: “chosen” financial fragility limits CB’s MoPo
- With AI strategic interaction/game becomes **lopsided** $\Rightarrow$ AI agents outsmart CB

4b. Financial Liquidity Provision & Macroprudential Policy

- Lending facilities and asset purchases
- Moral hazard:
 - Commitment not to bail out is not credible $\Rightarrow$ ever fine-tuned ex-ante regulation
(SVB 2023, ...)
- AI market participants **set traps**
& **“engineer” situations** to maximize bail-out
 - CB cannot understand traps ex-ante (like Move 37 in Game Go)
 - CB cannot verify **collusion** among AI traders
 - **Speed asymmetry**: regulated entity mutates faster than the regulator can adapt

4. Shifts in Fundamental Tradeoffs

- Under asymmetric understanding, *game* btw. CB and market becomes *lopsided*. For the authorities, *fine-tuning* a strategic game against a collectively smarter counterparty is a *losing strategy*.

⇒ **Four (counter intuitive) shifts** to allow CB & regulators to stay in the game:

1. Rules vs. discretion: from fine-tuning to blunt, robust rules
2. Transparency vs. opacity: end of constructive ambiguity (no hidden rules)
communicating two 2 audiences: humans, AI
3. Acting through markets vs. side-stepping them: (EMDE analogy)
direct channels as threat points
4. Aligning AI from within influence AI foundation models (reward function),
one generation ahead rule

Roadmap

1. Intro: Trust
2. From Asymmetric Information to **Asymmetric Understanding**
3. Financial Markets with AI Agents
 - in the Age of AI
 - Monetary Policy as a LoSegmentation, ...
4. Central Banking lopsided Game btw Central Bank and Market Participants
5. **Transition, Resilience,
AI Dependency and Power Concentration ⇒ systemic risk**
6. Conclusion

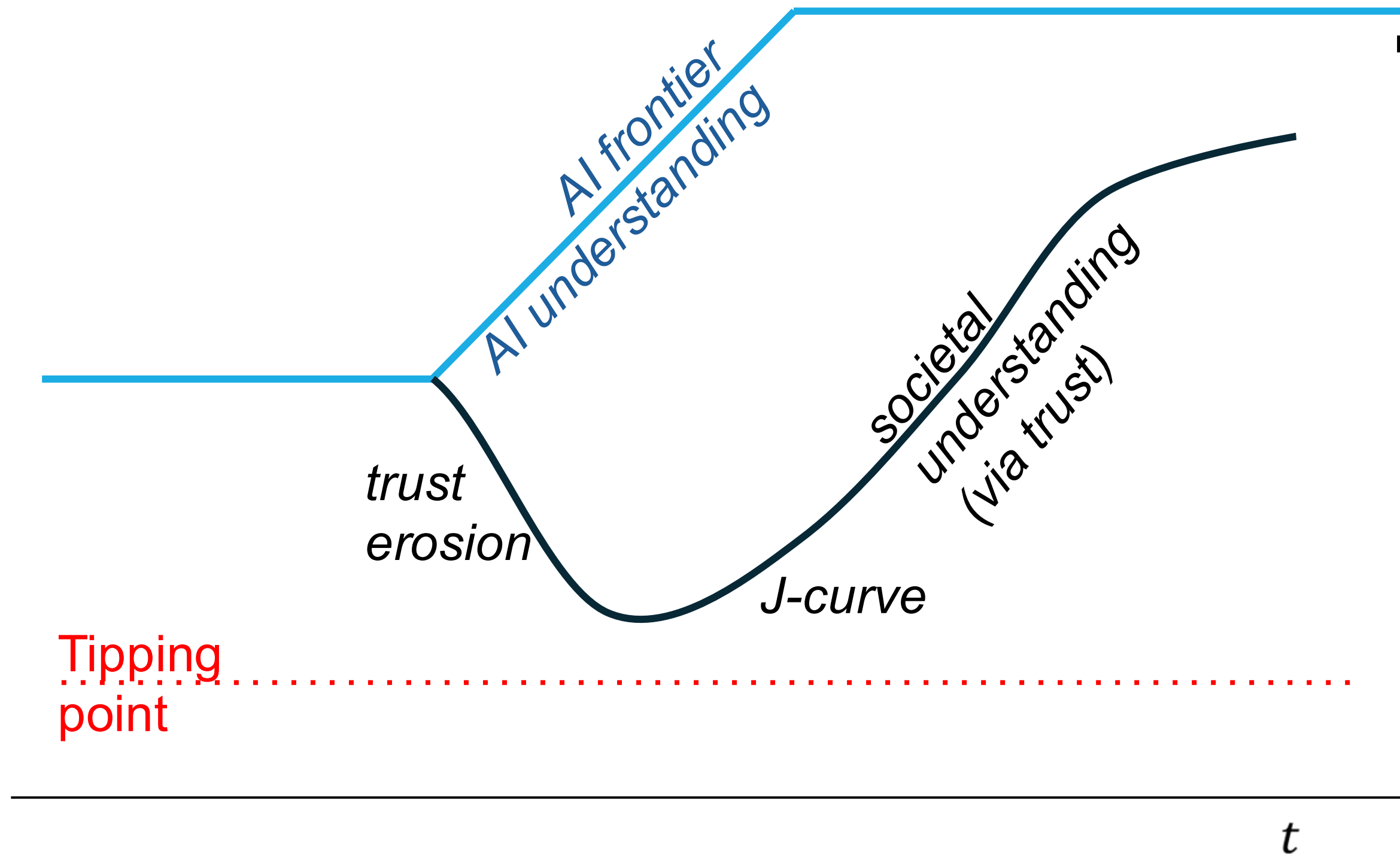
Erosion of Trust and Societal Understanding

- **Societal** understanding via **trust**

- Institutional trust arrangement $\Rightarrow$ it is enough to trust someone who understands

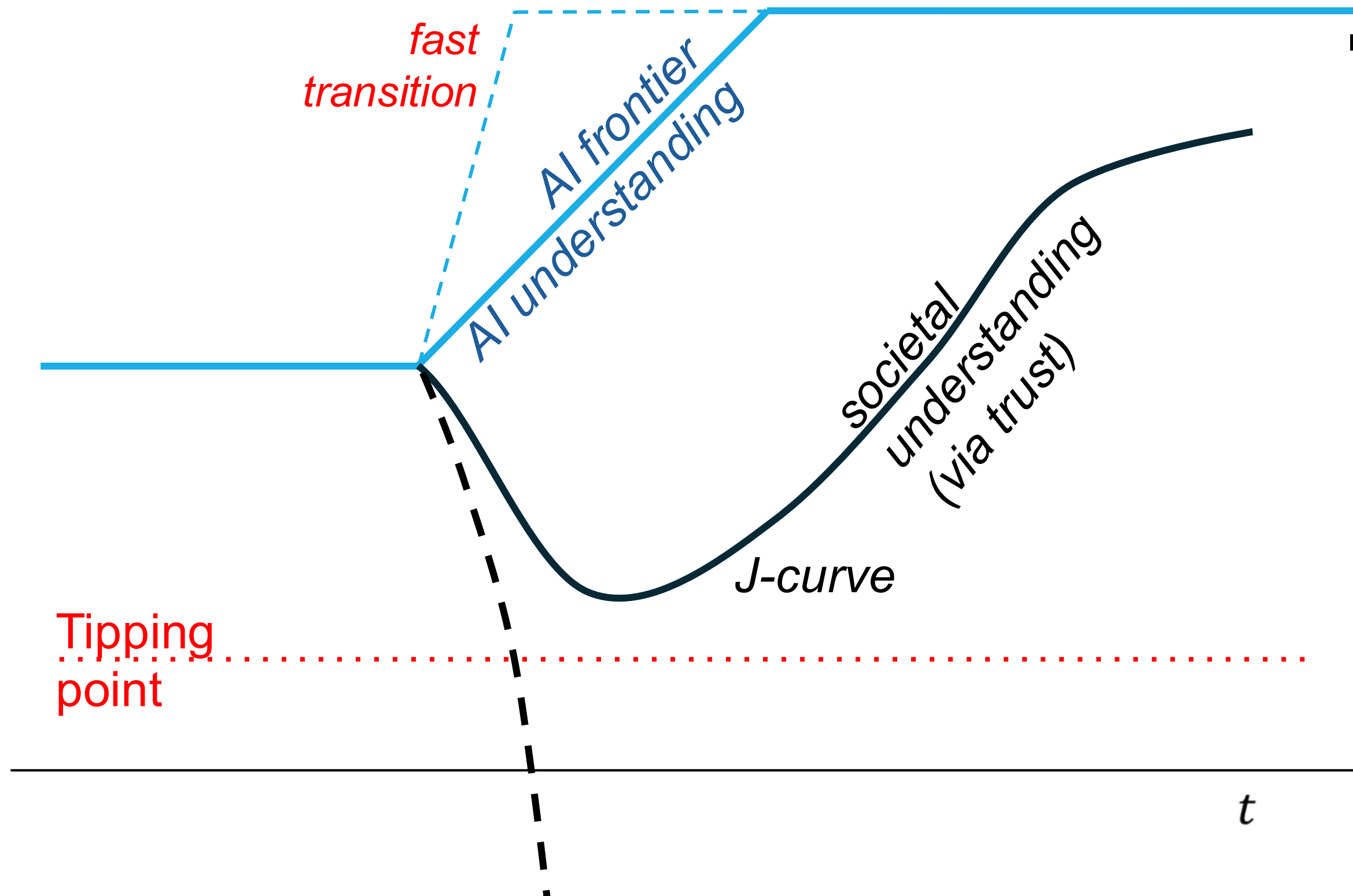


5. Transition Dynamics and Resilience



- **Transition** Phase: Setbacks and Risks $\Rightarrow$ resilience
 - Gap btw innovation and understanding
 - Repricing risks: moat disruption

5. Transition Dynamics and Resilience



- **Transition** Phase: Setbacks and Risks $\Rightarrow$ resilience
 - Gap btw innovation and understanding
 - Repricing risks: moat disruption

5. Tech Dependency, AI Addiction & Concentration of Power

- **AI Dependency**

- E.g. Mythos withdrawal June 2026
- Foreign dimension

⇒ systemic risk (“loss of control”)

- **Concentration of Power**

- Interoperability Regulation:
Lower switching costs across versions and vendors
 - Value creation vs. value capture

⇒ Stress tests with tech-dependency, AI market structure – kill switch illusion

6. Conclusion

*We live in a time when
“philosophical questions
become practical ones”.*

- AI introduce radical form of **Asymmetric Understanding**

- **Disrupts** current institutions & erodes **trust** (arrangement) at least initial setback
- AI agents have **different/deeper representation of the world**

1. **Financial Markets** with interacting AI agents

- EMH weaker since trading functions are non-explainable, collusion, pump & dump, flash crashes

2. **Monetary Policy** as Game btw CB and AI agents

3. **Liquidity Policy (LLR)** and Financial Regulation

⇒ strategic interaction btw CB and market becomes **lopsided**

- **Four shifts** in central banking trade-offs

1. Rules vs. discretion: from fine-tuning to blunt rules
2. Transparency vs. opacity: end of constructive ambiguity
3. Rather than recruiting markets, side-step them
4. Aligning AI from within influence AI foundation models (reward function)

- Transition resilience, tech-dependency & concentration of power ⇒ **new systemic risk**