# Eduard Talamas

# Introducing AI Agents

On Friday, January 17, Eduard Talamas joined Markus' Academy for a conversation on "Introducing AI Agents." Eduard Talamas is an Assistant Professor of Economics at the IESE Business School and a CEPR Research Fellow.

A few highlights from the discussion.[1]

- **A summary in four bullets**
    - AI agents are likely to be one of the key trends in AI this year. In the talk Talamas presented his recent paper with Enrique Ide ([2024](#)) on the topic
    - The paper builds on the canonical model from the organizational and knowledge economy literature (Garicano, [2000](#)), introducing AI a form of knowledge that can be scaled and, unlike human knowledge, is not constrained by time
    - Its key result is that the degree of AI autonomy will be critical in determining its winners and losers. Autonomous AI will benefit the most knowledgeable workers, while non-autonomous AI will do the opposite
    - The model may reconcile the conflicting empirical findings on whether AI benefits the least or the most knowledgeable workers

- [**[0:00](#)**] **Markus' introduction**
    - How is introducing a population of AI agents different from simply increasing the human population?
    - There are several types of AI agents: (1) Behavioral agents act based on simple or model-based rules, (2) Goal-based agents use algorithms to maximize a function (e.g. utility), (3) Learning agents are able to improve their behavior through experience, (4) Hierarchical agents handle complex tasks by breaking them into subtasks, (5) Multi-agent systems involve multiple agents interacting within a shared environment, often with different objectives
    - Economists traditionally model IT innovations as expanding the span of control. Will AI be similar?

- [**[04:47](#)**] **AI agents will likely arrive this year**
    - [Sam Altman](#) and [Dario Amodei](#) have recently highlighted AI agents as a key trend this year, with autonomy being their defining feature —the ability to independently pursue projects
    - Indeed, economists like [Gilian Hadfield](#) have urged the discipline to view AI not just as a tool but also as a novel economic agent capable of things like entering into contracts, setting prices, or hiring
    - As Jason Harline has highlighted, economists should leverage their models of human behavior to understand the impact of AI agents

---

[1] Summary produced by Pablo Balsinde (PhD student, Stockholm School of Economics)

- In their recent paper, Ide and Talamas ([2024](#)) try to build on the organizational and knowledge economy literature to argue that a population of AI agents is fundamentally different from a human population, and that their economic impact will crucially depend on their autonomy
- In their review of this literature, Garicano and Rossi Hansberg ([2015](#)) highlight that the use of a human's knowledge is constrained by their time, and organizations serve to make the best possible use of the available knowledge
- In contrast, the main insight of Ide and Talamas ([2024](#)) is that AI allows knowledge to be used at scale, releasing these time constraints and leading to a dramatic reorganization of firms

- **[[18:50](#)] Pre-AI knowledge economics**
  - Before considering AI, it is helpful to understand the canonical model of the knowledge economy literature (Garicano, [2000](#))
  - In it, there is a unit mass of humans, each with one unit of time and an exogenous level of knowledge. Competitive firms hire workers and organize production, where each worker uses their time and knowledge to complete a single project
  - (As a side note, in a separate paper, Ide and Talamas ([2024b](#)) consider AI in a setting where projects require multi-dimensional knowledge)
  - Although ex-ante projects are identical, during their completion problems arise with a random difficulty. A worker will complete the project (produce output) if its amount of knowledge is larger than the difficulty it faced
  - In organizing production firms can build a hierarchy, and place a high-knowledge worker as a "solver/manager." If a normal worker cannot solve a problem, they can ask their manager for help at the cost of the manager's time
  - The main idea then is that hierarchies allow the firm to shield managers from routine work, allowing them to specialize in exceptionally hard problems
  - As Alfred Sloan, the former head of General Motors, explained: "We do not do much routine work with details. They never get up to us. I work fairly hard, but on exceptions"
  - Two types of firms emerge: those with a single worker and layer and those with a two-tier structure composed of workers and managers
  - Further, occupational stratification emerges, where a knowledge threshold will determine whether someone is hired as a normal worker or a manager
  - Without hierarchies workers earn their expected output (their knowledge), so the wage schedule is just the 45 degree line. However, with hierarchies workers can see their wages improve, as workers earn the benefits of the improved allocation of knowledge to tasks

- **[[29:15](#)] Introducing AI into the canonical model**
  - In their model, Ide and Talamas try to study AI in a market for white collar labor and focus on three key properties of AI: (1) that in contrast with human knowledge it can be used at scale, (2) that it is a general purpose technology, and that (3) it can automate knowledge work
  - They introduce AI firms which own an exogenous amount of computing power and turn it into an exogenous level of knowledge (which amounts to how

advanced the AI is). Whether (and how many) workers lie above or below the knowledge level of AI will drive many of the results
- The AI firms are competitive and maximize profits (issues around oligopolies by AI firms are abstracted away)
- The paper draws a key contrast between autonomous and non-autonomous AI. Developers think of autonomy as an AI being able to execute a project; in the model an AI agent being autonomous means that it can pursue projects as a coworker (until it runs into problems it cannot solve and asks for help from the manager)
- A non-autonomous AI agent can only act as a solver/manager/copilot and provide advice when asked for help by workers encountering difficulties
- With hierarchies and AI five types of firms can emerge:



*Z, $Z_{AI}$ and S denote the knowledge of workers, AI agents, and managers*
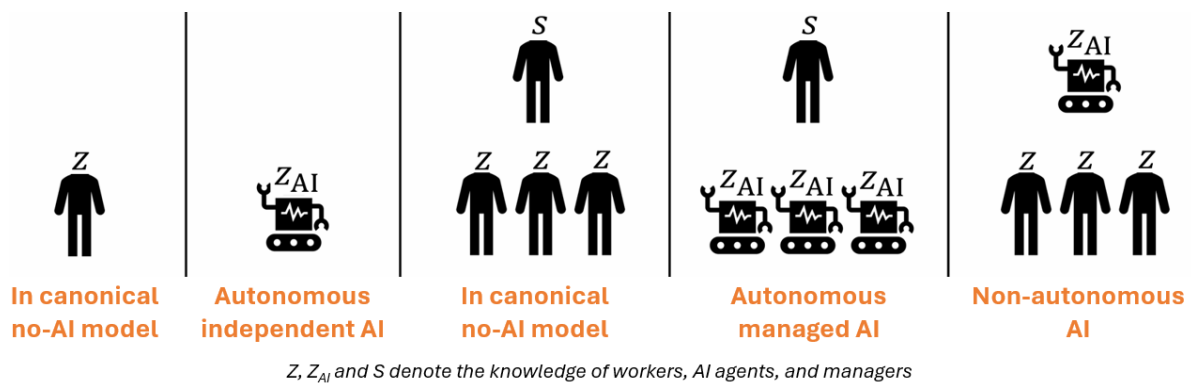
- It is worth noting that it is not possible to have an all-AI hierarchical firm. In contrast to workers, which have heterogeneous levels of knowledge, all AI agents (or units of compute) have the same fixed level knowledge; so if the bottom layer cannot solve a problem the manager AI will not be able to either
- In the benchmark model the amount of computing power is large relative to the amount of human time, so there will be some independent production by AI agents. This will pin down the rental rate of compute as the expected output of an AI agent (and by substitution it will also pin down the wage of workers with exactly the same knowledge as the AI)

- **[39:42] Effects of introducing (autonomous) AI agents**
  - Suppose a relatively basic AI with more knowledge than roughly a quarter of workers is introduced. Because its knowledge is below the threshold that turns workers into managers it will be used only autonomously (unable to act as a solver/manager)
  - Wages for a large share of workers, including many with more knowledge than the AI, will decline; these workers will have been substituted away
  - Interestingly, many of the more knowledgeable workers become managers, as the introduction of an AI that can be used as a worker increases the demand for problem solving
  - Those that were managers before the introduction of AI will see their wages grow. They will largely use AI in their hierarchies, and will be able to leverage their own knowledge more effectively

○ A smaller share of human workers (denoted by *I* below) will become independent producers. These are the workers who are not knowledgeable enough to turn into managers
○ The results are summarized in the chart below. The **red line depicts the wage schedule under a no-AI market**, and the **blue line the wage schedule with the introduction of an autonomous AI**:
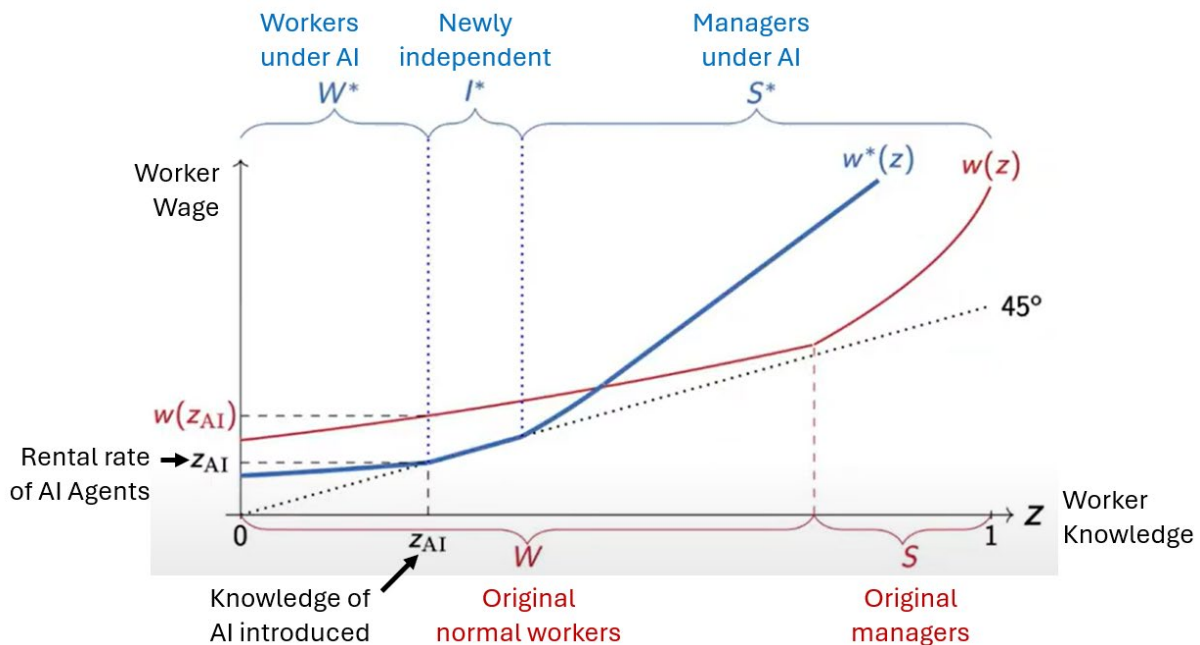


○ What happens if the AI introduced is more knowledgeable (higher $Z_{AI}$)? The general result is that total labor income goes up, and there are always winners at the top
○ The original managers with a still higher knowledge than the AI are unambiguously better off. The rest of the original managers will see their wages decline. Interestingly, the workers at the bottom will be slightly better off (see page 25 of the paper)
○ This is because as AI agents improve they will begin solving/managing, so workers will not be substituted as much. In effect, these least knowledgeable humans will have access to a good manager that is relatively cheap
○ To generalize: if $Z_{AI}$ is in the pre-AI worker region (see the **W red bracket**) the cutoff to become a solver/manager will come down. If $Z_{AI}$ falls in the pre-AI manager region the opposite will happen, and humans will move to routine work

● **[48:08] Effects of introducing (non-autonomous) AI agents**
○ Suppose through regulation it is prohibited for AI agents to pursue projects – that is, be in the bottom layer of hierarchies
○ In most cases the least knowledgeable will see their wages grow, and the most knowledgeable will see their wages decline. Intuitively, who benefits from an AI agent which is only helpful when one gets stuck? Those who get stuck a lot

- ○ Unambiguously, aggregate output will decline after the ban. This introduces a tradeoff between income inequality and output
  - ○ In the model the loss is incurred by the owners of compute. If these firms are in developed countries this introduces an international dimension to the tradeoff (Ide and Talamas, work in progress)
  - ○ As a side note, what makes AI agents different from general computers in the model is their autonomy and ability to solve problems. The literature had modeled the ICT revolution as a reduction in the communication costs between workers and managers (Garicano and Rossi-Hansberg, 2006)

- **[56:03] Conclusion**
  - ○ In summary, autonomous AI will benefit the most knowledgeable workers, while non-autonomous AI will benefit the least knowledgeable
  - ○ With this point, the model reconciles conflicting findings in the literature on whether AI benefits the least or most knowledgeable (e.g. Brynjolfsson et al. (2023) and Toner-Rodgers (2024))
  - ○ It also highlights that in future work researchers should not only collect data on AI's IQs and EQs, but also on Action Quotients
  - ○ Lastly, it points to the tradeoff policymakers face in regulating AI autonomy between labor income inequality and output

**Timestamps:**
**[0:00] Markus' introduction**
**[04:47] AI agents will likely arrive this year**
**[18:50] Pre-AI knowledge economics**
**[29:15] Introducing AI into the canonical model**
**[39:42] Effects of introducing (autonomous) AI agents**
**[48:08] Effects of introducing (non-autonomous) AI agents**
**[56:03] Conclusion**