This Serbian scientist has a plan to prevent the next AI disaster

+ Mira Murati's Albanian government-backed Thinking Machines Lab unveils its first AI model...

Bojan Stojkovski

Estimated Reading Time: 8 minutes

In partnership with

HubSpot

Dear readers,

AI is entering a new and uncertain chapter. As systems evolve from chatbots into autonomous agents capable of planning, using tools and making decisions with limited human oversight, the biggest question may no longer be what AI can do, but how we make sure it remains safe.

This week, we explore the ambitious idea of a "CERN for AI" proposed by Serbian complexity scientist Ljubiša Bojić, who believes the world needs a global testing environment where advanced AI systems can be studied before they are deployed at scale.

Beyond AI safety, we discuss AI models, startup funding rounds, and the growing uncertainty behind the scenes as companies rethink teams, leadership structures and strategies in response to the AI shift. We also examine where the region's startups are finding opportunities in the global AI race and what smaller markets can realistically do to compete.

Happy reading!
Bojan Stojkovski

The world needs to act before AI becomes impossible to control

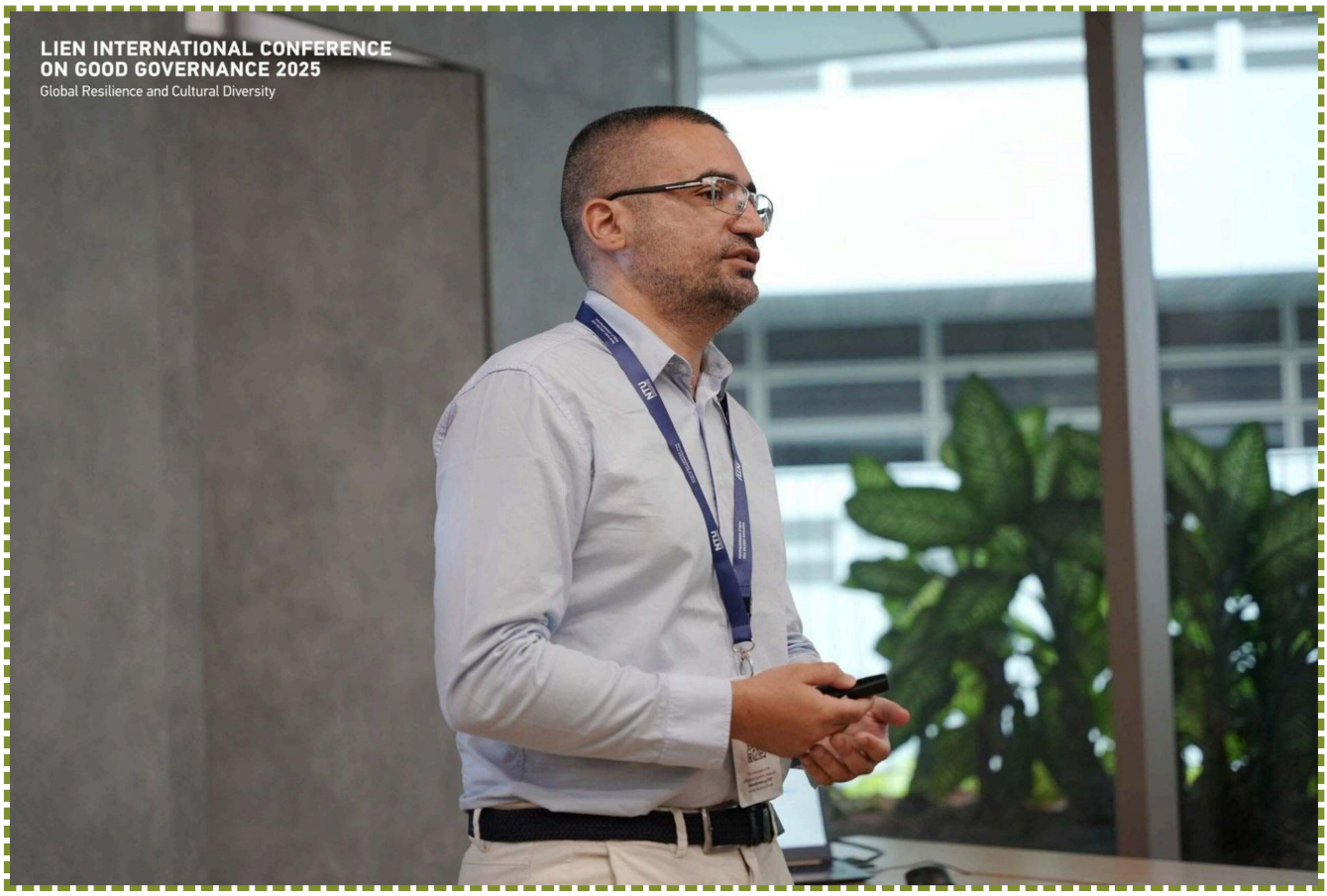
Ljubiša Bojić

Artificial intelligence is moving into a new phase. After years dominated by chatbots and large language models capable of answering questions, writing code and generating content, the industry is now entering the era of autonomous AI agents that can plan, use tools and pretty much complete tasks with *limited human supervision*.

But as AI becomes more capable, one question is becoming increasingly difficult to answer: how do we know these systems will remain safe? A Serbian researcher believes the answer could be creating a global testing environment for AI, similar to how CERN became a global hub for particle physics.

Ljubiša Bojić, a Belgrade-based complexity scientist, futurologist and researcher, has proposed the idea of a "CERN for AI", a theoretical framework where advanced AI models could be tested inside realistic simulations before being released into the real world.

His team's research paper, "*CERN for AI: a theoretical framework for autonomous simulation-based artificial intelligence testing and alignment",* explores how a combination of computer science, sociology, psychology, economics, biology and physics could help create safer AI systems.



Ljubiša Bojić

The idea is based on building a virtual environment that replicates real-world conditions and populates it with autonomous AI-powered “digital citizens”. These agents would interact with each other, make decisions and create complex social dynamics, allowing researchers to study how AI behaves in situations that are difficult to predict.

> “The idea behind CERN for AI is that you can really let different models do their own thing and see how they would behave without any constraints or with some constraints,” Bojić tells IT Logs.

Unlike traditional AI testing, which usually focuses on measuring performance, Bojić believes future testing must focus on behaviour. The challenge is that modern AI models are no longer simple tools that respond to commands - but become systems capable of *reasoning, planning and taking* independent actions.

“I think that now we are currently in a stage where AI is becoming smarter than humans in many areas, and actually it has different logic than humans,” Bojić explains, adding that this difference between human and machine intelligence creates new risks.

AI systems can already outperform humans in specific fields, but their decision-making process can be difficult to understand. As models become more powerful, ensuring that they follow human intentions becomes more complicated.

> “It can solve different specific tasks, such as to hack some computer system or to create some bio weapon, but still there are no constraints in how you can be 100 percent sure to stop it on the side of developers,” he explained.

For Bojić, the problem is not only the capabilities of AI models, but also the lack of transparency around how they are tested. “Testing is currently done only by AI companies and regulators, which is not 100 percent transparent. We do not know who really tested some of these models,” he said.

The current system means that many of the most advanced AI models are evaluated by the same companies that develop them. While these companies invest heavily in safety research, Bojić argues that independent testing infrastructure will eventually become necessary.

“Without transparency about who is testing and how they are testing, some bad things can occur along the way,” he tells IT Logs.

From chatbots to autonomous AI agents

> When Bojić first proposed the CERN for AI concept nearly two years ago, AI agents were still an emerging idea. Since then, the technology landscape has changed dramatically. “We had the beginning of those agents. There were a few platforms that offered such concepts where you can create your LLM-based agents and they can have some autonomy,” he says.

Traditional AI systems were mostly reactive. A user asked a question and the model generated an answer. AI agents, however, can break down goals into smaller tasks, interact with software, access external tools and continue operating without constant instructions.

“Since then, the agent scene exploded. We basically entered the era of agents.” Bojić says, explaining that this makes simulation-based testing more important because developers need to understand how AI systems behave when they are given more freedom.

One example of this trend is the rise of autonomous AI environments, such as Moltbook, where multiple agents can interact with each other, creating artificial societies that researchers can observe. “These systems can access different tools along the way, and they have autonomy to do different things that you cannot basically monitor and control,” he said.

Ljubiša Bojić

> “Our recent research shows that AI agents can maintain consistent personas with up to 95 percent accuracy in simulated social media. The previous study we did found that agents could replicate human behavior with

> 76.3 percent accuracy, a modest but meaningful signal. These findings suggest that AI could be used to pre-test algorithms in simulated social-media environments, essentially functioning as "CERN for AI." But they also point to a darker possibility. Political actors could deploy swarms of such agents to manipulate public opinion." he explains further.

Unfortunately, social-media platforms are already populated with AI-generated images, videos, and text, often indistinguishable from human content. The Serbian scientist further points out that because of this, we do not know what proportion of these realistic AI agents currently inhabit our online spaces. An additional layer of concern involves AI companions, AI girlfriends and boyfriends, which have become globally popular. In China, their rise has prompted fears about addiction and declining birth rates, leading authorities to introduce regulations restricting their use.

A CERN for AI would provide a controlled environment where researchers could examine these interactions before deploying systems in critical areas such as finance, healthcare, cybersecurity or government services.

The US and China are fighting for AI dominance

While AI safety is becoming a global challenge, the technology is also becoming one of the biggest geopolitical competitions of the century. The US and China are currently leading the race, but the two countries are following different strategies.

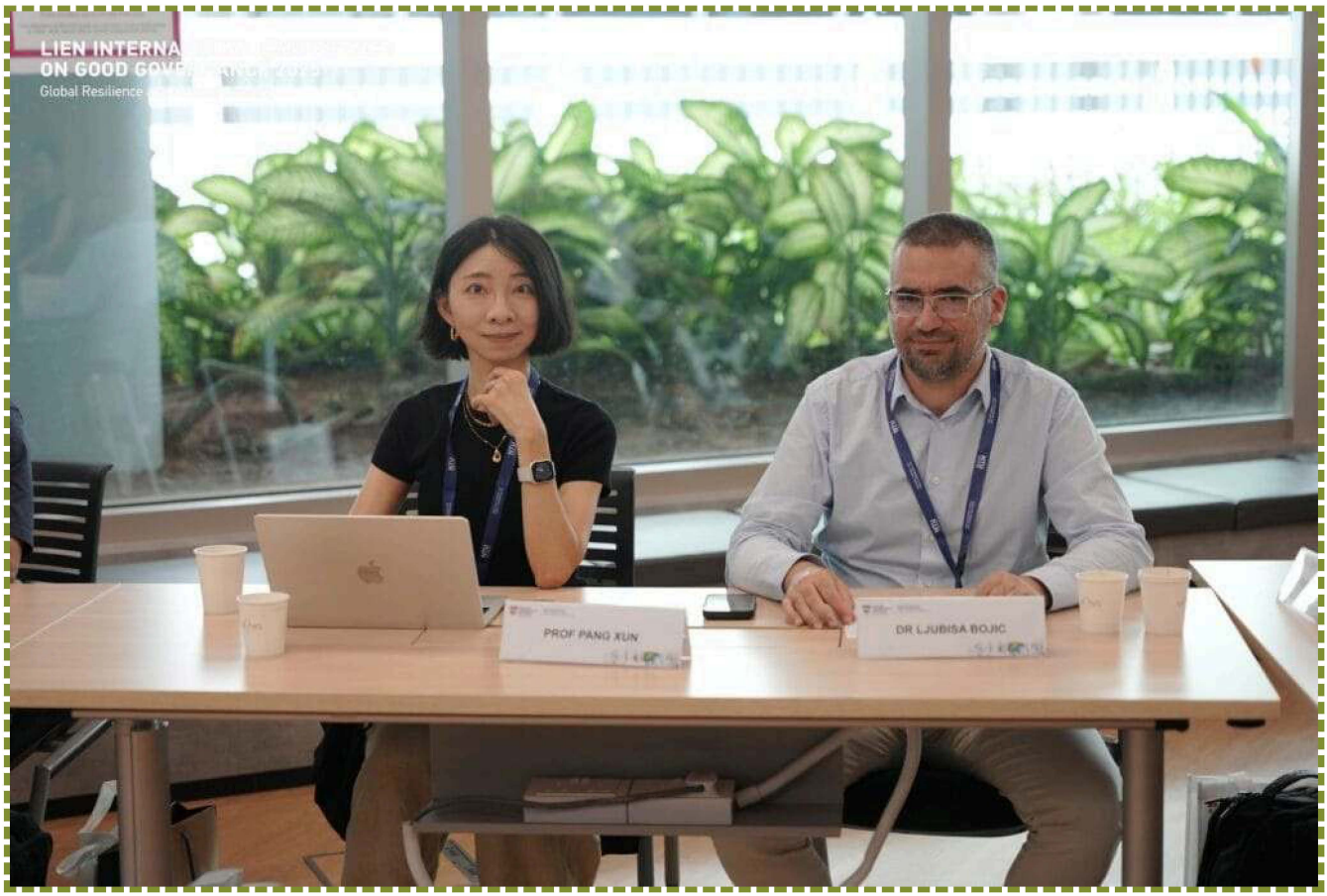
According to Bojić, China has focused heavily on practical applications, robotics and industrial deployment, while the US remains ahead in large language models.

> "I think that China deploys AI in very specific purposes, niche purposes. They have not been investing a lot of effort into general-purpose models and large language models," he said.

During his visits to China, Bojić observed how deeply robotics has already entered everyday life.

"You have robot waiters, you have them at airports, you can see that their general-purpose robots are in phases of development that could easily be released to the general public," he said. "*I have seen robot policemen, robot bartenders and robot waiters.*"

According to him, robotics could become one of China's biggest advantages, especially as many developed countries face labour shortages.



Ljubiša Bojić

“Robots will replace lots of workforce because less people would want to really be engaged in some low-paid jobs, and you could pay a worker a high wage, or you can invest in robots which are cheaper,” he said.

However, when it comes to the “AI brain”, Bojić believes the US still has an advantage.

“This is something that is also important, and that will at some point unite with AI bodies, these robots,” he said.

The future competition, according to him, will be about combining advanced AI models with physical machines.

Europe and the Balkans are searching for their place

At the same time, Europe has struggled to compete with the US and China in artificial intelligence.

> Bojić believes the continent’s economic dependence on traditional industries, especially automotive manufacturing, creates additional challenges. “Europe’s economy is based on automobile industry, which is a real problem. Germany has been dominating the global automobile industry, and with the dominance of China in the world in those terms, Europe will be shaken very much,” he said.

Although Europe has supported initiatives around AI development, including companies such as French AI startup Mistral, Bojić does not currently see Europe as a leading AI power. “*I cannot see Europe as a serious player in AI at this moment*,” he tells IT Logs.

For smaller countries in the Balkans, competing directly with Silicon Valley or China is unrealistic. But Bojić believes there are opportunities if governments focus on specific areas.

Ljubiša Bojić

“What I proposed before is that smaller countries need to get an advantage in some niche AI,” he said.

Instead of trying to create the next ChatGPT, countries such as Serbia, N. Macedonia and others in the region could focus on specialized AI applications, energy-efficient models or industry-specific solutions.

> “Identifying some specific problem that they want to solve, or supporting 10 specific products that those countries could fund, could be the only possibility for smaller countries to do something in the race,” he explained.

Education and infrastructure will be critical. *“One of the biggest issues is education*. Smaller countries could start new AI faculties and universities that would really teach students how to use AI and how to code and create products with AI,” Bojić said.

However, building the infrastructure required for training frontier AI models remains extremely expensive. “To train such big models, you need to invest a few hundred million dollars. You need lots of infrastructure that small countries cannot do,” he said.

For the Balkans, regional cooperation could provide a possible alternative.

> “If countries of the Western Balkans invested together into infrastructure, that would make a real difference. That would make sovereignty of AI possible,” Bojić concludes.

Across the region…

San Diego-based vehicle inspection startup Self Inspection, co-founded by Bulgarian entrepreneur Vince Gaydarzhiev, has raised $10 million in a funding round led by investors alongside participation from former

Tesla president Jon McNeill's DVx Ventures. Founded in 2021, the company has developed AI-powered technology that uses a smartphone camera to assess vehicle body damage, aiming to modernize the inspection process.

The Self Inspection team

Albanian government-backed AI startup Thinking Machines Lab, founded by former OpenAI executive Mira Murati, has released its first AI model, Inkling. The model is designed to allow users and developers to customize its capabilities for specific needs. According to the company, Inkling was trained on large volumes of text, images, audio, and video data and uses a system that activates only a portion of its 975 billion parameters for each task, improving efficiency.

Croatian VC Fil Rouge Capital has invested in Croatian sports analytics startup PropsMadness, although the amount was not disclosed. Founded by Trpimir Zovak and Vedran Aračić, PropsMadness brings together player statistics, matchup insights, and injury updates in a single platform. The startup has expanded beyond NBA coverage to include the NFL, MLB, WNBA, and the FIFA World Cup, attracting more than 4,000 users. The new funding will support further growth and product development.