



Human-Centric Trusted AI for Data-Driven Economy

Masugi Inoue¹ and Hideyuki Tokuda²

National Institute of Information and Communications Technology

inoue@nict.go.jp

1, Director, International Research Advancement Office, Global Alliance Department

2, President



国立研究開発法人

情報通信研究機構

National Institute of Information and
Communications Technology

The sole national research institute in the field of ICT in Japan

Three roles:

- Research Institute, Research Funding Agency, Public Service Agency

4th 5-year-Term from April 2016 to March 2021

Personnel: ~ 1100


- Researchers: ~ 530
- PhDs: ~ 460

Public Services:

- Japan Standard Time
- Space Weather Forecast
- Wireless Equipment Testing & Calibration
- Cybersecurity Training



EU-Japan Joint Calls funded by EC and NICT

		ICT-2013-EU-J d)	Extending the cloud paradigm to the Internet of Things – Connected object and sensor clouds within the service perspective		
		FP7	Call 1	ICT-2013-EU-J e)	Federation of testbeds : Control, tools and experiments
				ICT-2013-EU-J f)	Green & content centric networks
H2020	Call 2	EUJ-3-2014	Access networks for densely located users		
		EUJ-4-2014	Experimentation and development on federated Japan - EU testbeds		
	Call 3	EUJ-02-2016	IoT/Cloud/Big Data platforms in social application contexts		
		EUJ-03-2016	Experimental testbeds on Information-Centric Networking		
		SC1-PM14-2016	EU-Japan cooperation on Novel ICT Robotics based solutions for active and healthy ageing at home or in care facilities		
	Call 4	EUJ-01-2018	Advanced technologies (Security/Cloud/IoT/BigData) for a hyper-connected society in the context of Smart City		
		EUJ-02-2018	5G and beyond		

We need to identify future topics for Joint Calls in 2020 and later. 3

We Need Trusted AI

Prof. Amnon Shashua, SVP Intel, CEO & CTO of Mobileye

<https://newsroom.intel.com/editorials/experience-counts-particularly-safety-critical-areas/>

Experience counts, particularly in safety-critical areas.

Now Is the Time for Substantive Conversations about Safety for Autonomous Vehicles



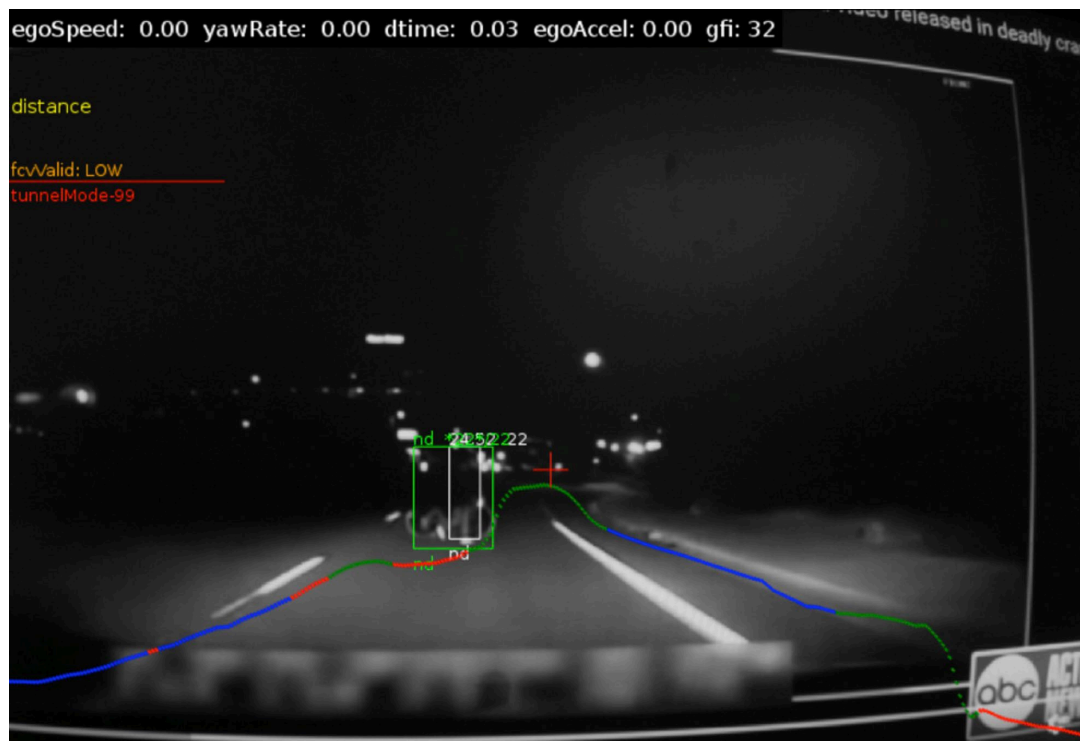
By Prof. Amnon Shashua

Society expects autonomous vehicles to be held to a higher standard than human drivers. Following the tragic death of Elaine Herzberg after being hit last week by a self-driving Uber car operating in autonomous mode in Arizona, it feels like the right moment to make a few observations around the meaning of safety with respect to sensing and decision-making.

First, the challenge of interpreting sensor information. The video released by the police seems to demonstrate that even the most basic building block of an autonomous vehicle system, the ability to detect and classify objects, is a

Mobileye could detect a person ONE second before crash based on the release video data. He said...

1. Experience Counts
2. Transparency
3. Redundancy



Mobileye ADAS's response to the recorded video of the Uber crash

*Advanced driver assistance systems

Social Interaction



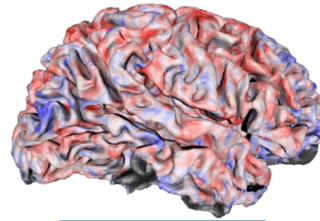
Emotion, stress, mental illness, chronic pain



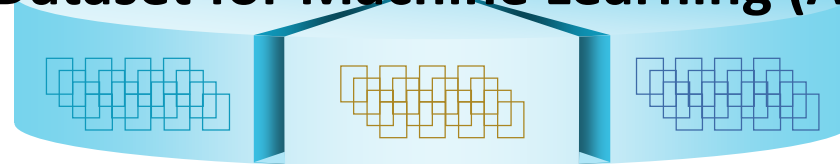
Physical performance



Large-scale brain information data



Dataset for Machine Learning (AI)



Applications

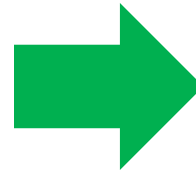
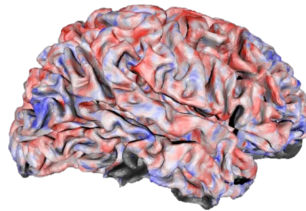
- Emotion evaluation
- Marketing
- Education
- Prevention/Treatment of psychiatric disorders and pain
- Biomarker
- Psycho-immunology
- Skill up such as sports

Reproducing What You See from Your Brain

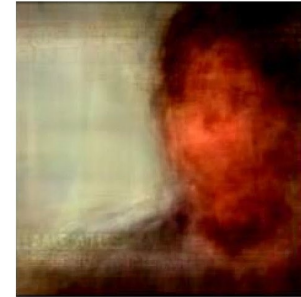
Presented Clip



Brain Activity



Reconstructed Image

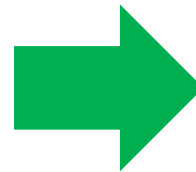
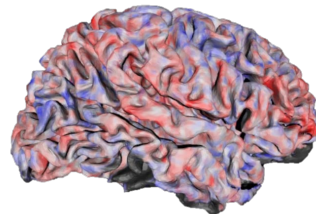


(A) Nishimoto, et al., *Current Biology*, 11th October 2011

Presented Clip



Brain Activity



Generated Sentence

A group of people standing on the beach.

(B) Matsuo, et al., *Proc. ACL SRW 2016*, 7th August 2016, arXiv, 19th January 2018

Motivation for Human-Centric Trusted AI

Present:

Modern data-driven applications using ML are getting popular in many application domains.

Spam filter, Credit card fraud detection, Face recognition, Recommendation, Stock trading, Medical diagnosis, etc

Future:

Needs safety, real-time, and interpretability of AI for critical social applications.

Transportation, Finance, Security, Medicine, etc



Human-centric and Trusted AI



Issues for Human-Centric Trusted AI

1. Black-box AI/Unexplainable AI
2. Integration with formal methods and control models
3. Necessary and sufficient data and testing
4. Real-World Security

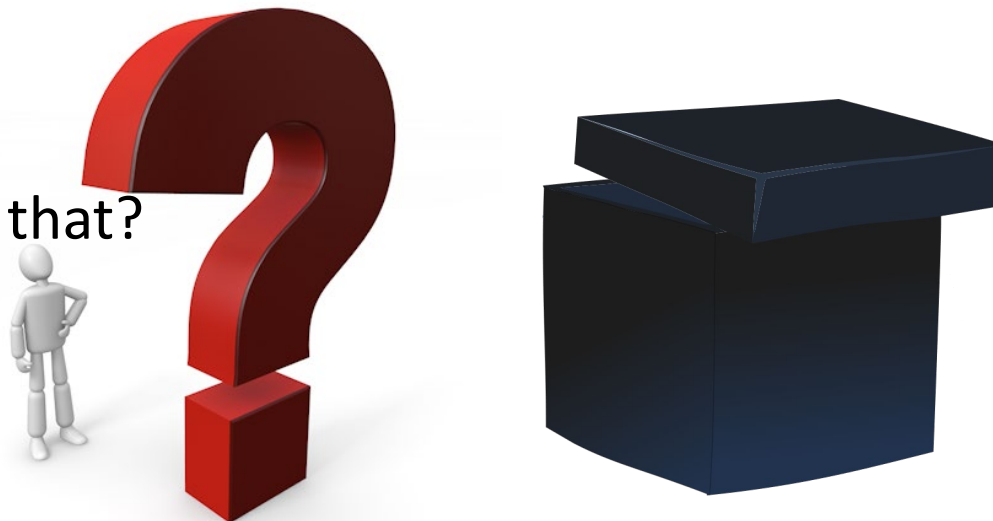
Black-box AI/Unexplainable AI

“Explaining how complex algorithms reach conclusions is critical as artificial intelligence becomes more deeply embedded in everyday life”

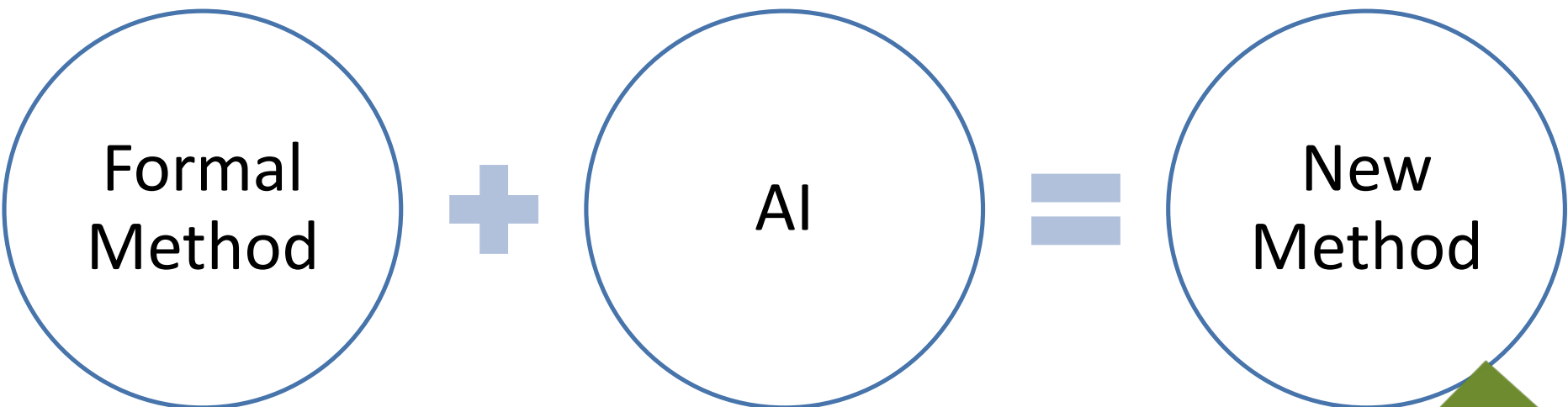
“If it’s finding patients that need special attention in the hospital, or wanting to know why your car stopped in the middle of the road, or why your drone turned around and didn’t do its mission ... then you really need an explanation,”

Source: WSJ “Inside Darpa’s Push to Make Artificial Intelligence Explain Itself”, Aug 10, 2017

Why did AI do that?



Integration of Traditional Formal Method and AI



Predictive

- Modeling complex systems
- Can verify properties
- Can ensure correct behavior mathematically

Unknown, Adaptive

- Modeling from data

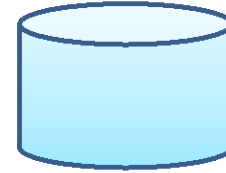
**Versatility,
Efficiency,
Accuracy**



Necessary and Sufficient Data and Testing

Necessary and Sufficient Data?

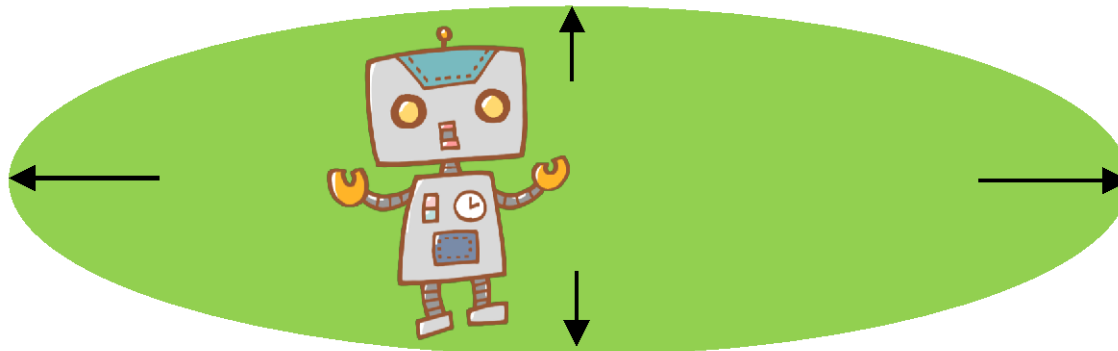
- Type
- Volume
- Variation
- Fairness (unbiased)



Necessary and Sufficient Testing?



Real World



Real-World Security

Not only cyber-world security but also protecting the real world from intentional attacks against AI is necessary.

Current AI misread “STOP”, recognized it as “SPEED LIMIT 45”.

