

06.12.2018

# AI ethics: A cornerstone of AI success

Riba Gábor – Senior Manager, PwC

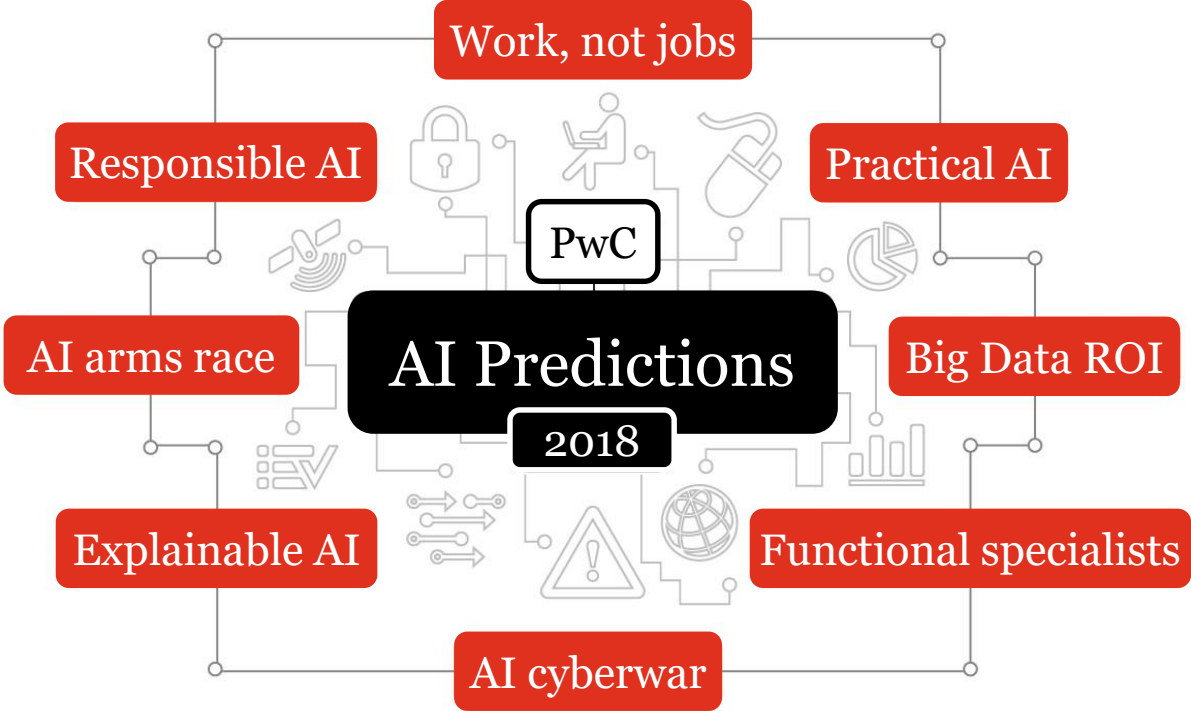


AI FORUM HUNGARY

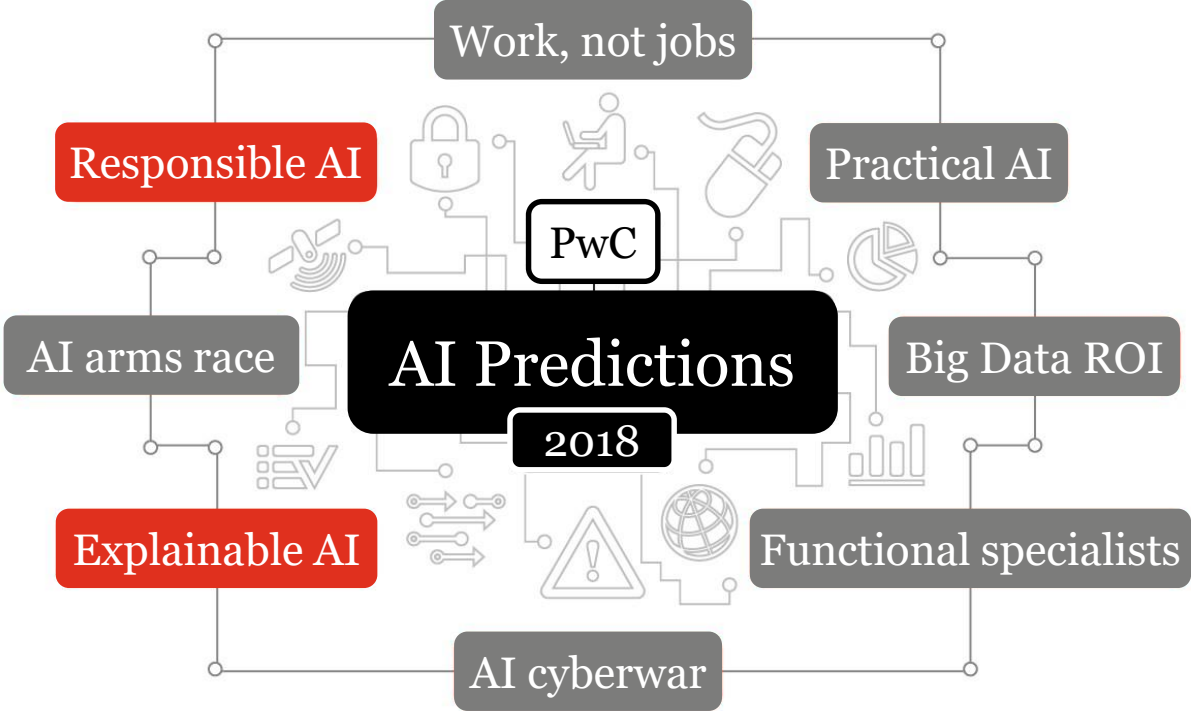
Section: Embedded  
ethics in AI design

[www.pwc.com/digital](http://www.pwc.com/digital)  
*Strictly private and confidential*

# PwC's AI predictions in 2018



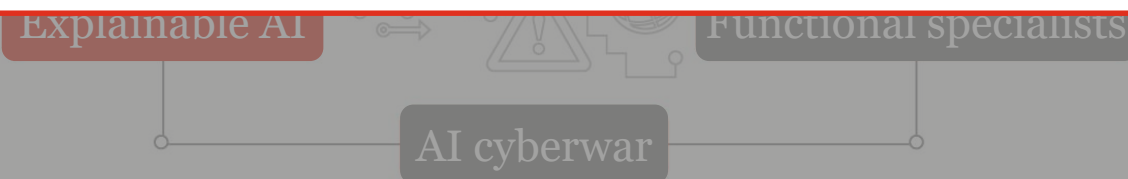
# Regarding AI ethics we need to focus on responsible and explainable AI



# Ethical AI consists of Explainable and Responsible AI



$$\text{Explainable AI} + \text{Responsible AI} = \text{Ethical AI}$$



# Responsible AI deployment should incorporate the concept of Explainable AI if a company aims for an ethic AI

## Explainable AI

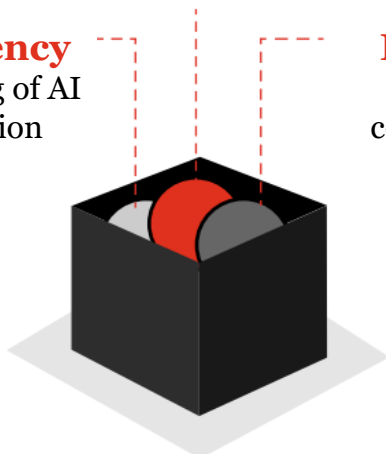
### What is inside the AI black box?

#### Explainability

Understanding reasoning behind each decision

#### Transparency

Understanding of AI model decision making

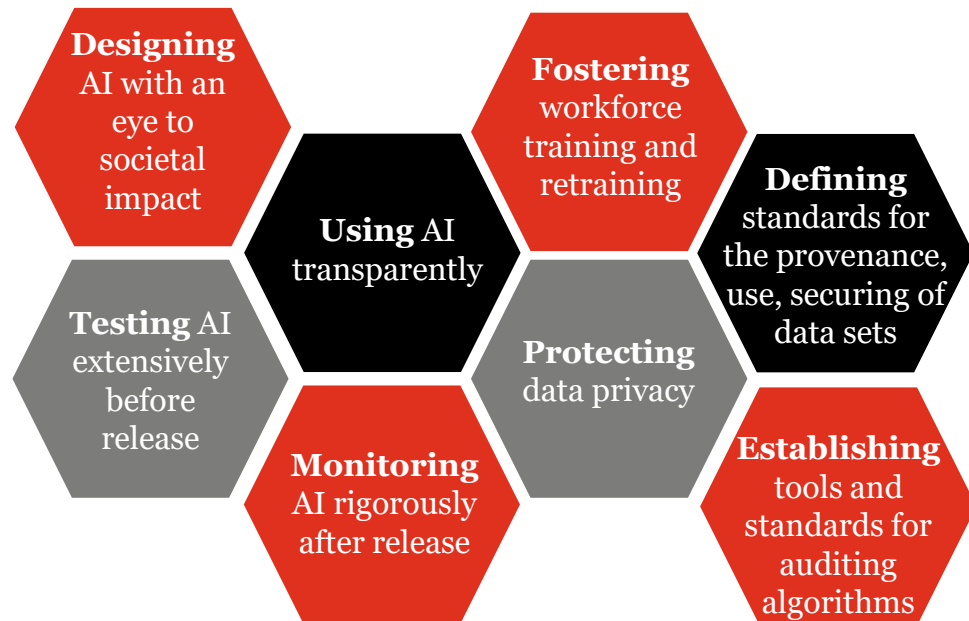


#### Provability

Mathematical certainty behind decision

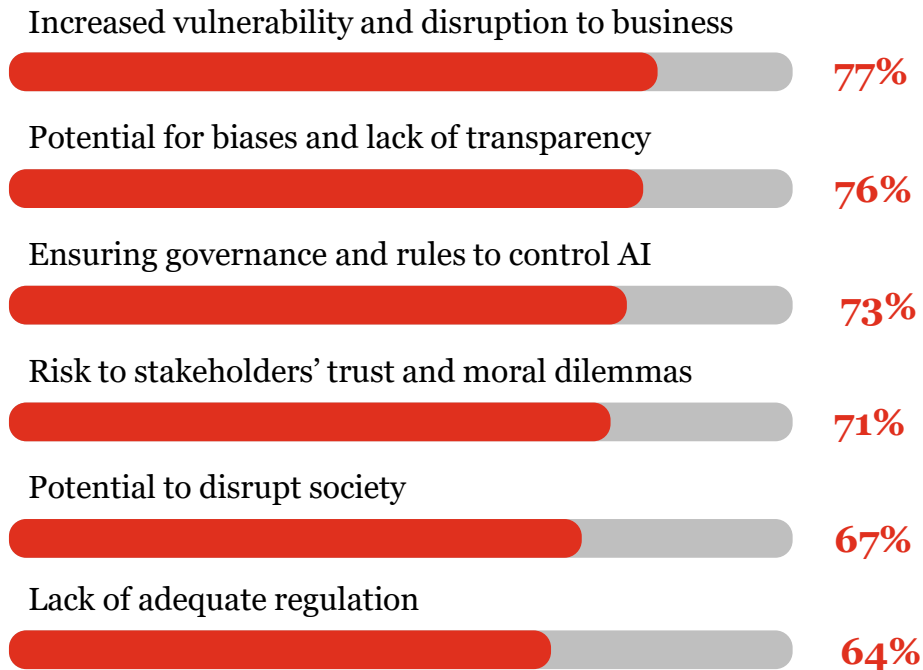
## Responsible AI

### What makes AI „responsible“?



# Enterprises mostly afraid of increased vulnerability and disruption to their business when adapting AI

## What's holding AI back in the enterprise?



The reasons that hold back enterprises from adapting AI are all show **a need for an ethical AI solution.**

# Who is responsible for providing ethical AI solutions?



**The following groups have all released sets of principles:** The World Economic Forum's Center for the Fourth Industrial Revolution, the IEEE, AI Now, The Partnership on AI, Future of Life, AI for Good, and DeepMind.



**Collaboration between public and private sector institutions**, as more governments explore the use of AI to distribute services efficiently, they're engaging citizens in the process. In the UK, for example, **the RSA** (Royal Society for the encouragement of Arts, Manufactures and Commerce) is **conducting a series of citizen juries on the use of AI** and ethics in criminal justice and democratic debate.

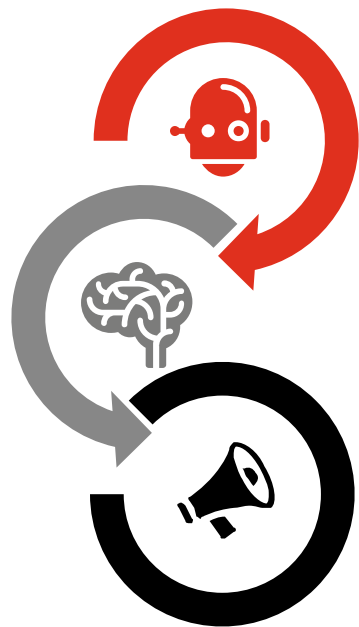


**Self-regulatory organizations (SROs)** may take the lead with responsible AI. An SRO **would bring users of AI together around certain principles**, then oversee and regulate compliance, levy fines as needed, and refer violations to regulators. It's a model that has worked in other industries.

The following use cases show why is it necessary to provide ethical AI solutions



# AI use case (1/4)– Microsoft Tay, the nazi chatbot



On March 23, 2016, **Microsoft released** a **‘teen girl’ chatbot** on Twitter named Tay for the purpose of engagement and entertainment.

Tay was **not able to truly understand the meaning of words** not to mention the context of the conversations.

She **promoted Nazism**, attacked feminists and Jewish people, and **denied** historical facts such as **the Holocaust**.



# Let's see how Amazon Echo communicates with Google Home



# AI use case (2/4)- Amazon Alexa, is she always listening to you?



## Who is Alexa?

Alexa, is **Amazon's virtual personal assistant**, it lives inside the company's Echo smart speaker, among other devices, and offers users the ability to dictate commands to the assistant to control products throughout their home, listen to music, and more.



A woman in Portland told, that her **Amazon Echo device had recorded a conversation** between her and her husband, **then shared it** with one of her husband's employees in Seattle

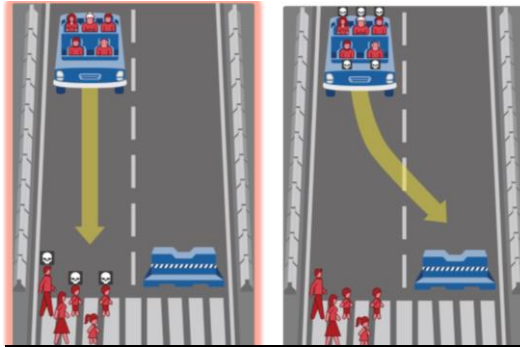


Users of Amazon's digital assistant, Alexa, have reported versions of the same unsettling event: **being startled** as they went about their day **by Alexa letting out an eerie laugh.**

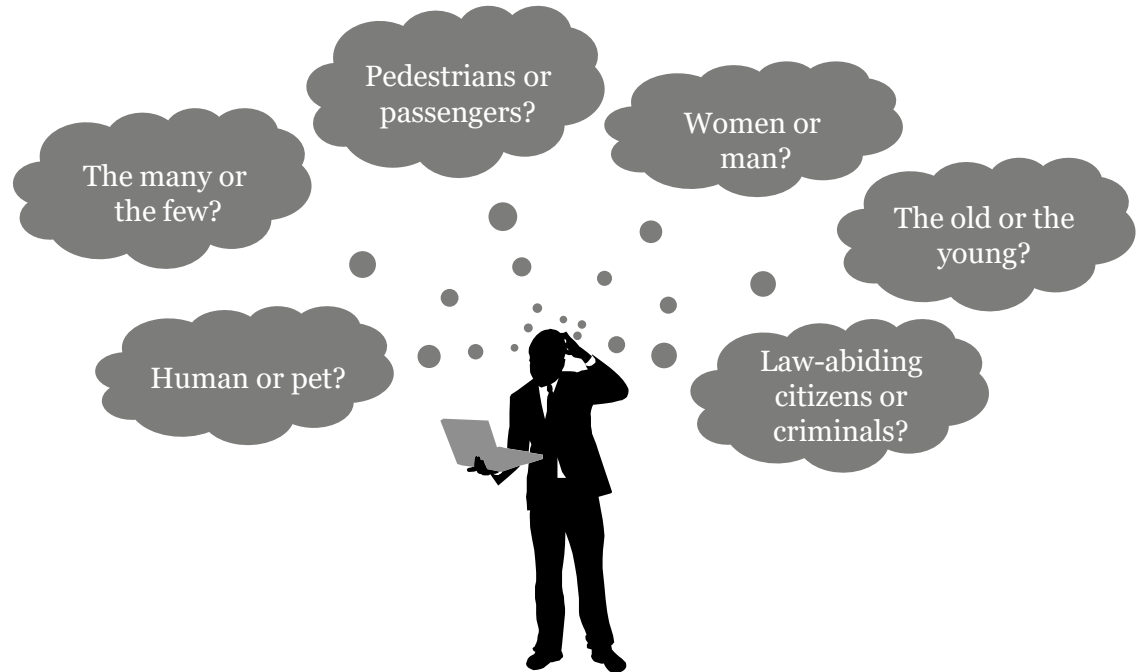


If the **user mentions** how much the user would **like to go to a restaurant** while on the phone, **a recommendation might be sent** while the user is still engaged in the conversation that enables the user **to make a reservation at the restaurant.**

# AI use case (3/4) – The Moral Machine, who would a self-driving car kill?



## How should you decide?



### What is Moral Machine?

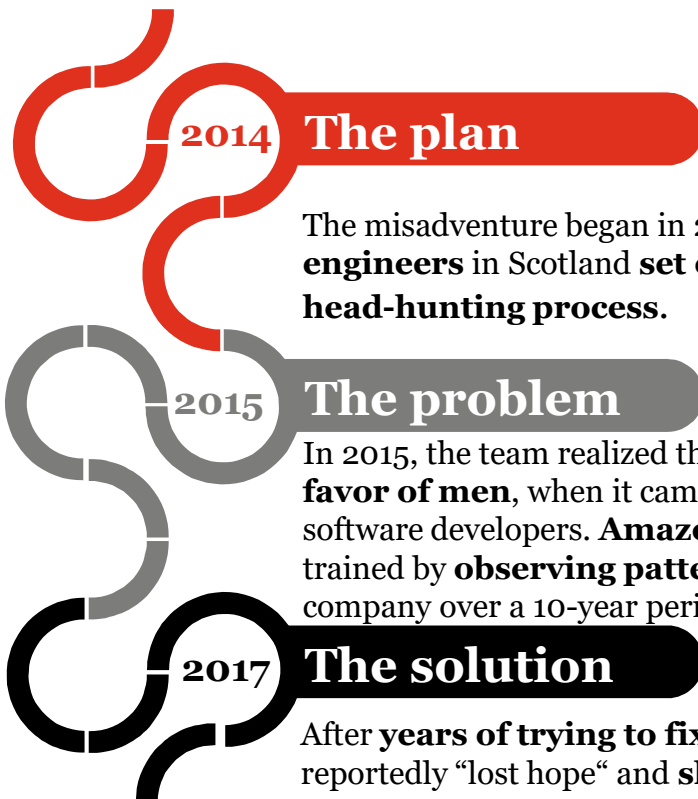
**Moral Machine** is a platform, that **shows you moral dilemmas**, where a driverless car must choose the lesser of two evils, such as killing two passengers or five pedestrians.

## AI use case (4/4) – The hiring AI tool which favored men



Amazon hiring tool

Its aim was **to mechanize** the company's **head-hunting process**, by creating a program that would scour the Internet for worthwhile job candidates.



2014

### The plan

The misadventure began in 2014, when **a group of Amazon engineers** in Scotland **set out to mechanize** the company's **head-hunting process**.

2015

### The problem

In 2015, the team realized that **its creation was biased in favor of men**, when it came to hiring technical talent, like software developers. **Amazon's computer models** were trained by **observing patterns in resumes** submitted to the company over a 10-year period.

2017

### The solution

After **years of trying to fix** the project, **Amazon** brass reportedly "lost hope" and **shuttered the effort** in 2017.

# Thank you!



**Gábor Riba**

*Senior Manager*

*PwC Hungary*

mobile: + 36 30 527 5398

[gabor.riba@pwc.com](mailto:gabor.riba@pwc.com)



© 2018 PricewaterhouseCoopers Hungary Ltd. All rights reserved. PwC refers to PricewaterhouseCoopers Hungary Ltd. and may sometimes refer to the PwC network. Each member firm is a separate legal entity. Please see <http://www.pwc.com/structure> for further details.

This publication has been prepared for general guidance on matters of interest only, and does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice. No representation or warranty (express or implied) is given as to the accuracy or completeness of the information contained in this publication, and, to the extent permitted by law, PricewaterhouseCoopers Hungary Ltd., its members, employees and agents do not accept or assume any liability, responsibility or duty of care for any consequences of you or anyone else acting, or refraining to act, in reliance on the information contained in this publication or for any decision based on it.