ObjectivAIze

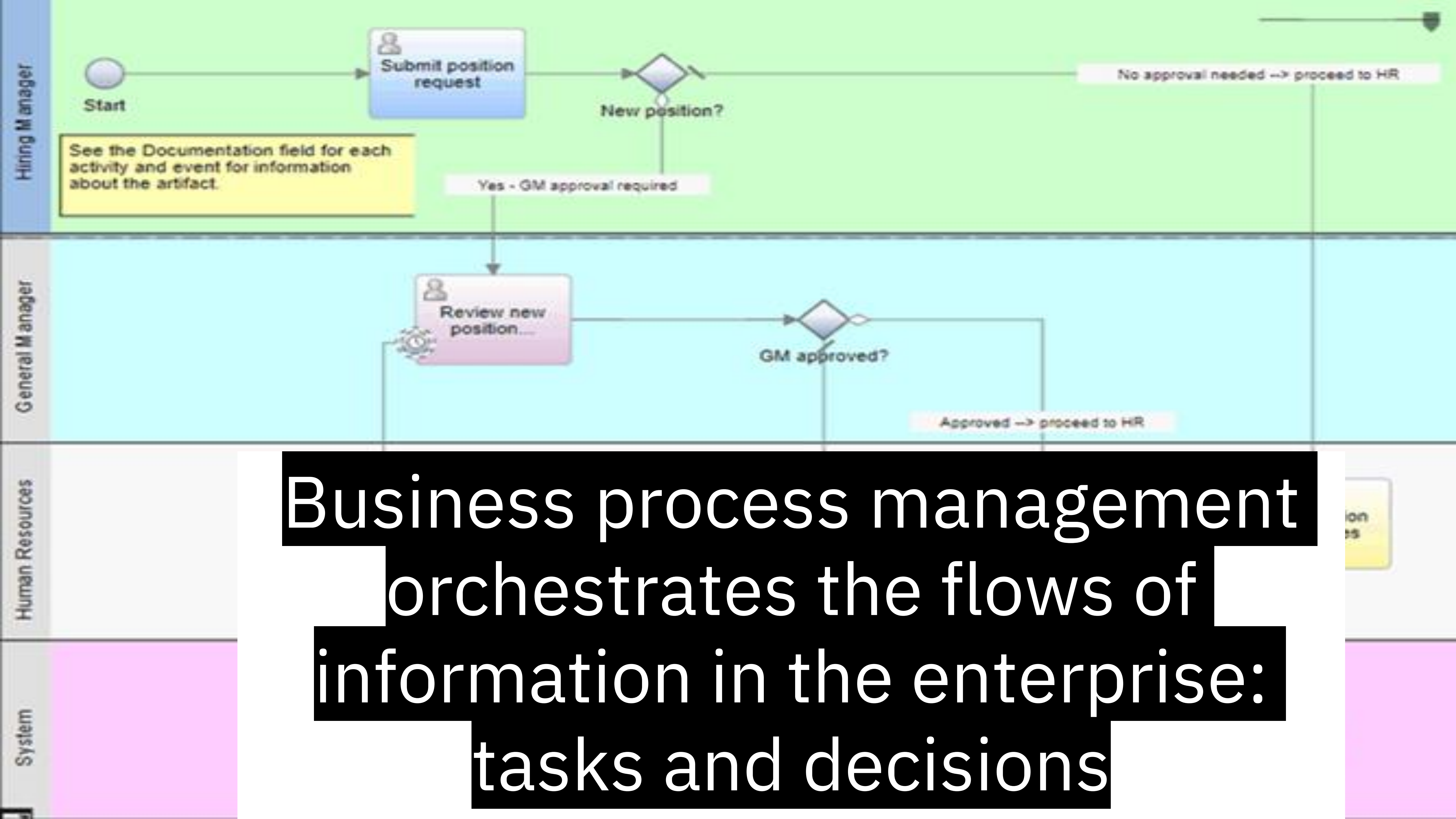
“View or interpret objectively without the influence of personal feelings or opinions »
Oxford dictionary

Mesurer performance et biais dans la décision augmentée, pour déterminer les conditions idéales de la collaboration humain-algorithme

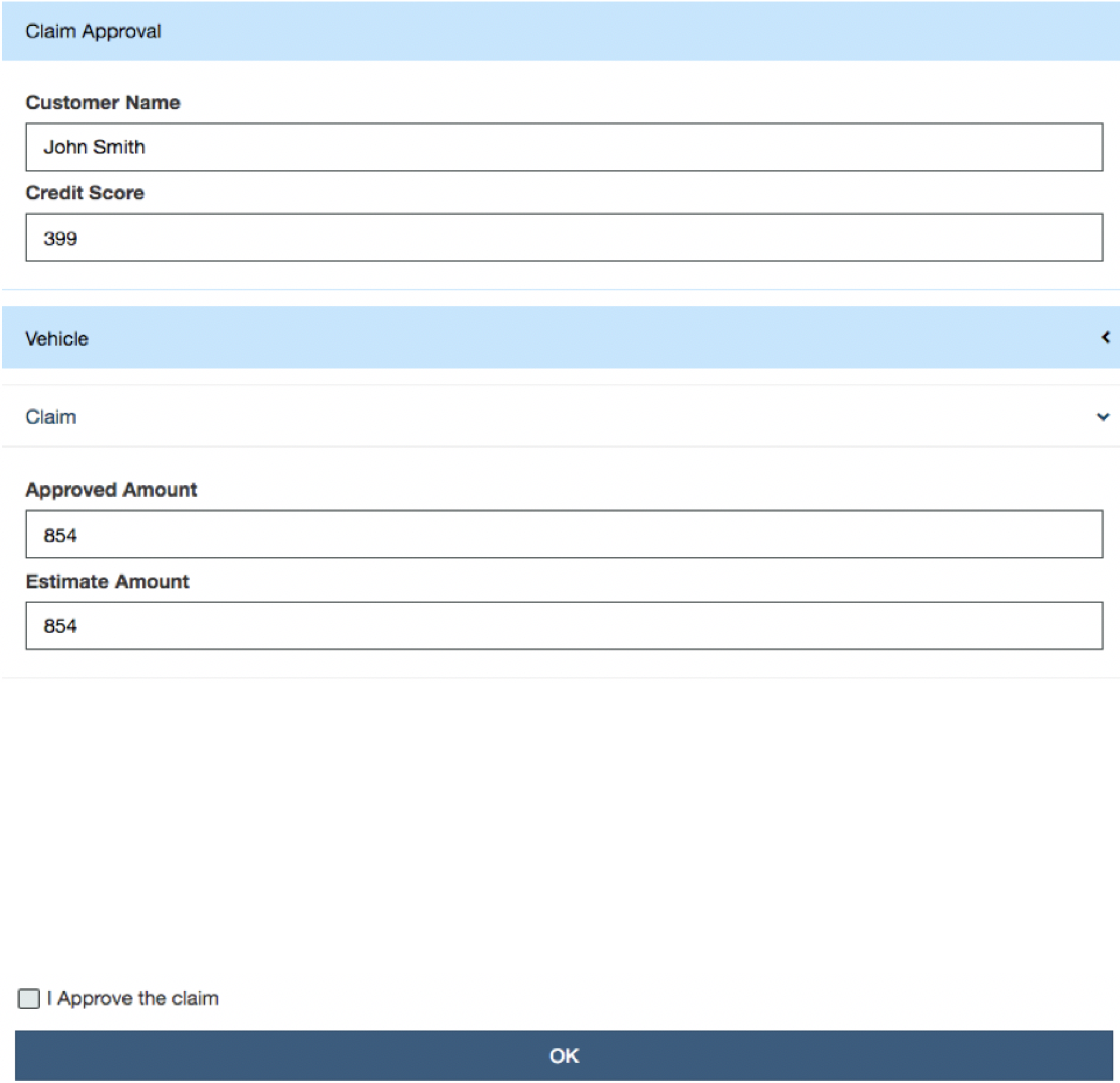
Oct. 7th, 2021

IBM

Business process management orchestrates the flows of information in the enterprise: tasks and decisions



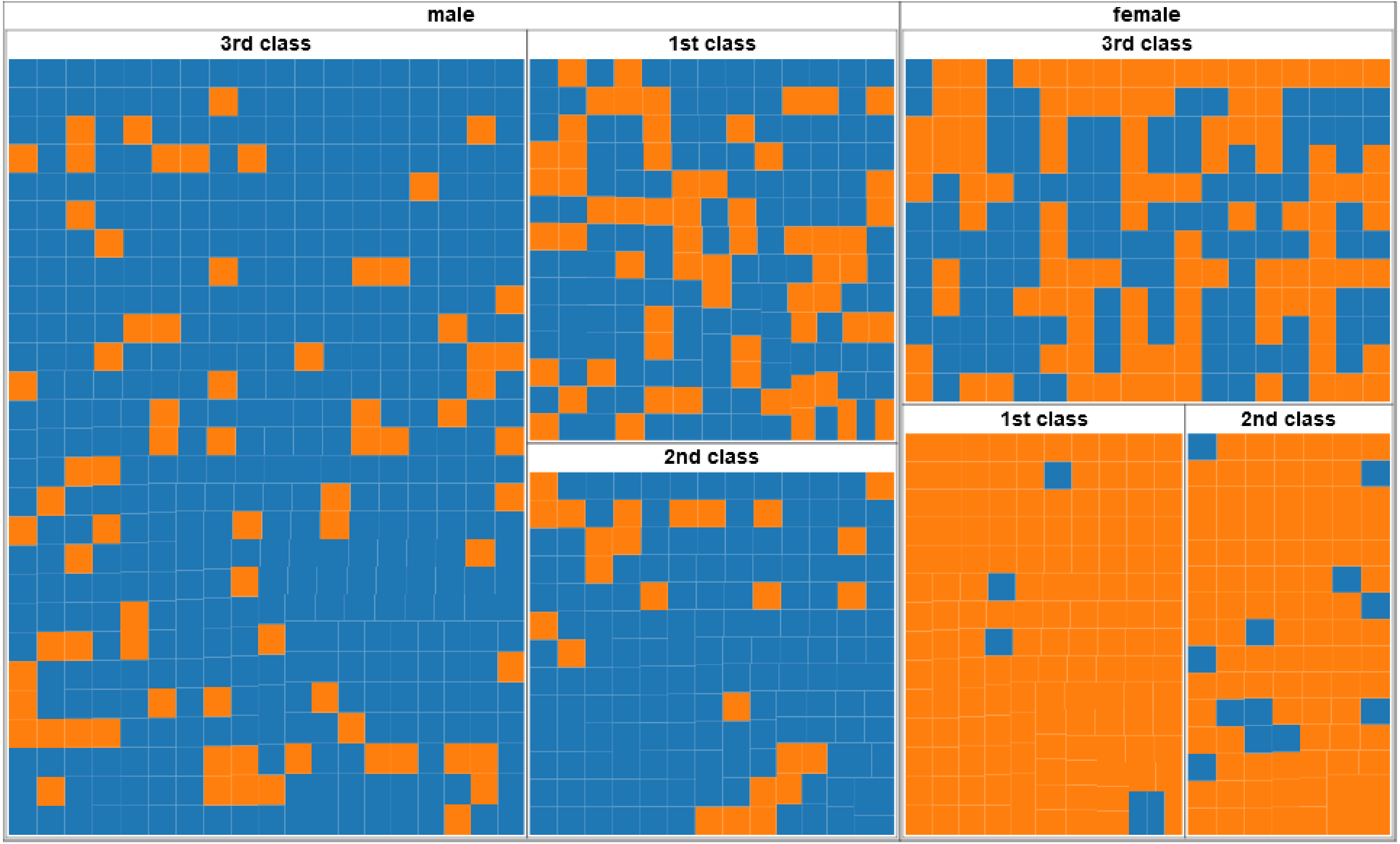
Statistical learning can help humans make better (more consistent) decisions



Augmented Decision Making

Algorithms can help humans make better decisions faster:

- Humans stay in charge, leverage context.
- Algorithms leverage past information, rules or statistical inference.


male
female
3rd class
1st class
2nd class
3rd class
1st class
2nd class

Passenger data:

Passenger n°:	179
Class aboard:	1
Sex:	female
Age:	49-64
Number of siblings or spouses aboard:	0
Number of parents or children aboard:	2
Fare:	31.0-
Embarkment area:	Cherbourg
Title:	Mrs

Make your decision here:

Survived

Died

3 / 20

Passenger data:

Passenger n°:	287
Class aboard:	1
Sex:	male
Age:	17-32
Number of siblings or spouses aboard:	1
Number of parents or children aboard:	0
Fare:	31.0-
Embarkment area:	Southampton
Title:	Mr

Success rate of the algorithm: 75%.
The algorithm recommends:

Died

Make your decision here:

Survived

Died

1 / 20

IBM Extreme Blue 2020. Powered by IBM Cloud

Passenger data:

Passenger n°:	45
Class aboard:	3
Sex:	male
Age:	17-32
Number of siblings or spouses aboard:	0
Number of parents or children aboard:	0
Fare:	07.9-14.5
Embarkment area:	Southampton
Title:	Mr

Success rate of the algorithm: 75%.
The algorithm recommends:

Survived

Make your decision here:

Survived

Died

4 / 20

Questions raised by Augmented Decision Systems (ADS)

What sort of ADS can be provided in Business processes?

- Decision trees
- Nearest neighbors
- Others (non-explainable)

Do ADS improve **accuracy** of decisions?

-> metrics of performance, both for the algorithm and the joint system

Do ADS introduce **automation biases**, or, on the contrary allow compensating algorithmic biases?

-> Measure biases and resistance.

Accountability transfer between human decision-maker and designer of the system

-> ethical dilemma, already explored in avionics and military systems.

We need **metrics** to address those questions, not just guidelines, recommendations and regulations

European regulation project on AI, Article 14 - Human oversight

1.**High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen** by natural persons during the period in which the AI system is in use.

2.Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter.

3.[…]

4.The measures referred to in paragraph 3 shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances:

(a) [transparency]

(b) **remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (‘automation bias’), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;**

(c) [explainability]

(d)**be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system;**

[…]

Related Art

Decision Theory

Rational decision theory vs. naturalistic decision theories.

Biases study (order effect, prompting...)

Risk vs. Uncertainty

Process control

Performance degrades when:

- The system is too bad (<70%)
- The system is too good (far superior to the human-> overreliance)

Recommender Systems

-> Algorithm aversion

Visual Analytics

-> perceptual effects

Industrial security & critical decision support (medical, avionics...)

Work process changes

=> Risk replaced by uncertainty: acceptability issues

Results (1) decision aid effectiveness

Control condition vs. recommendation:

M1 = 1.014, so the combination of human + algorithm does better than the human alone.

But not in the 80% case: $\mathbf{M_1 = 0.977}$

And, we would assume, not if success is < 70%

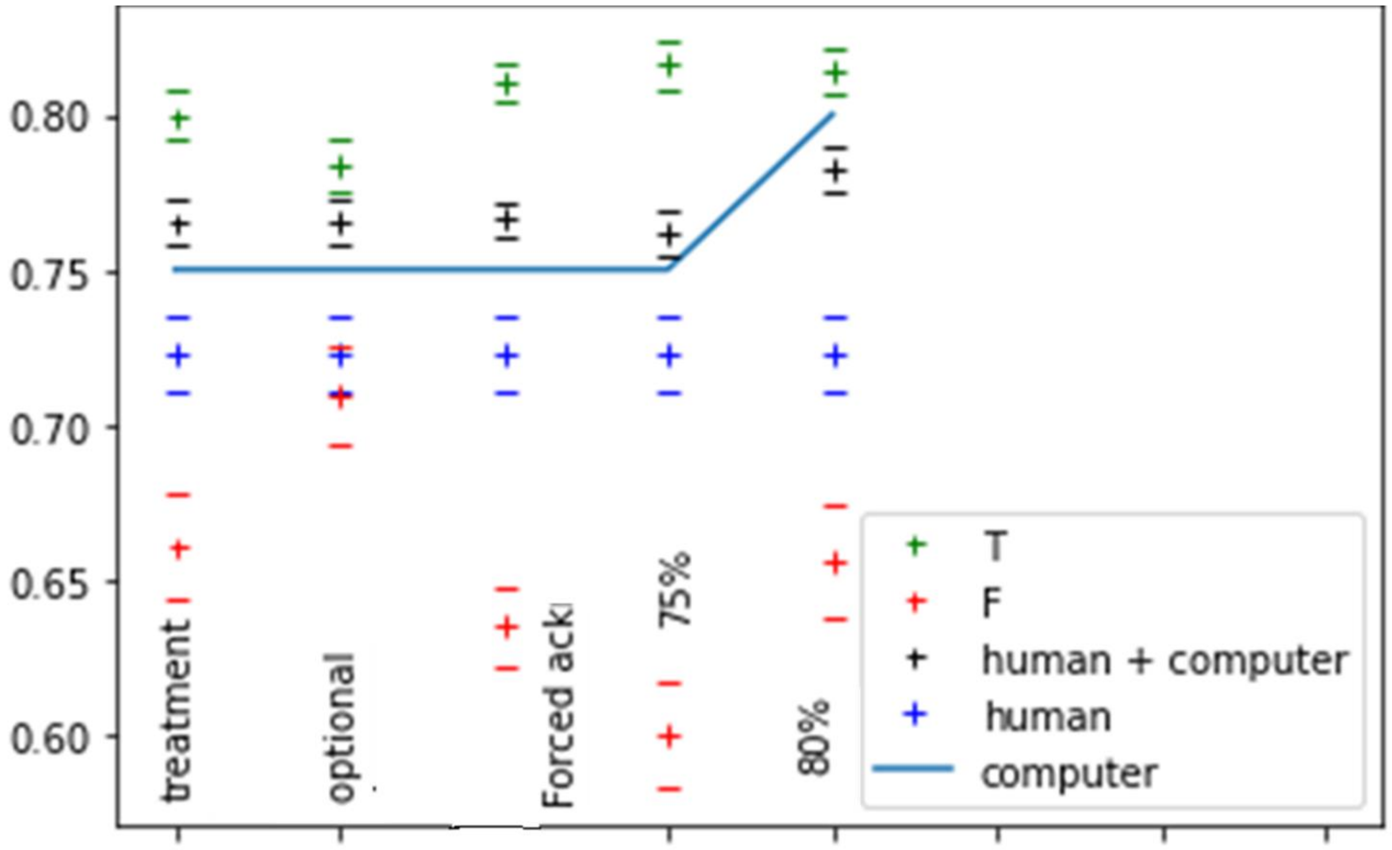
	coefficient	95% confidence interval
Control condition (human alone)	0.7230	[0.6948, 0.7512]
With decision aid	0.7604	[0.7530, 0.7682]
“Algorithm alone”	0.75	--

	coefficient	M_1
Control condition (human alone)	0.7230	0.9664
With decision aid (new run)	0.7651	1.020
Optional display	0.7655	1.020
Forced acknowledgment	0.7660	1.021
Reminder of 75%	0.7619	1.016

Results (2) Presentation influence

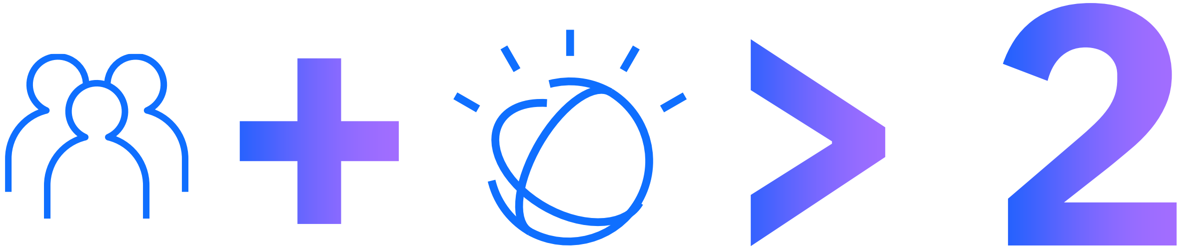
Forced acknowledgment maximizes the collaboration (raw performance)

Optional display maximizes the resistance (lowers algorithm influence, without compromising performance too much). -> when we want to minimize automation bias.



74% Of companies are exploring or deploying AI[1]

Supporting the idea that:



(1) IBM Global AI Adoption Index 2021

  6 October 2021 IBM confidential

But at the end of the day, **all our clients struggled with the same key questions:**

Who should decide?

Yes	No	No
Human	Algorithm	Human+ Algorithm

?

How is AI influencing Human decisions?

- Exercise of critical mind
- Automation bias
- Order or similarity bias
- Decision Fatigue
- Timing effect
- Expertise effect

Optional Display

Forced Display

Artificial Order

Similar Cases

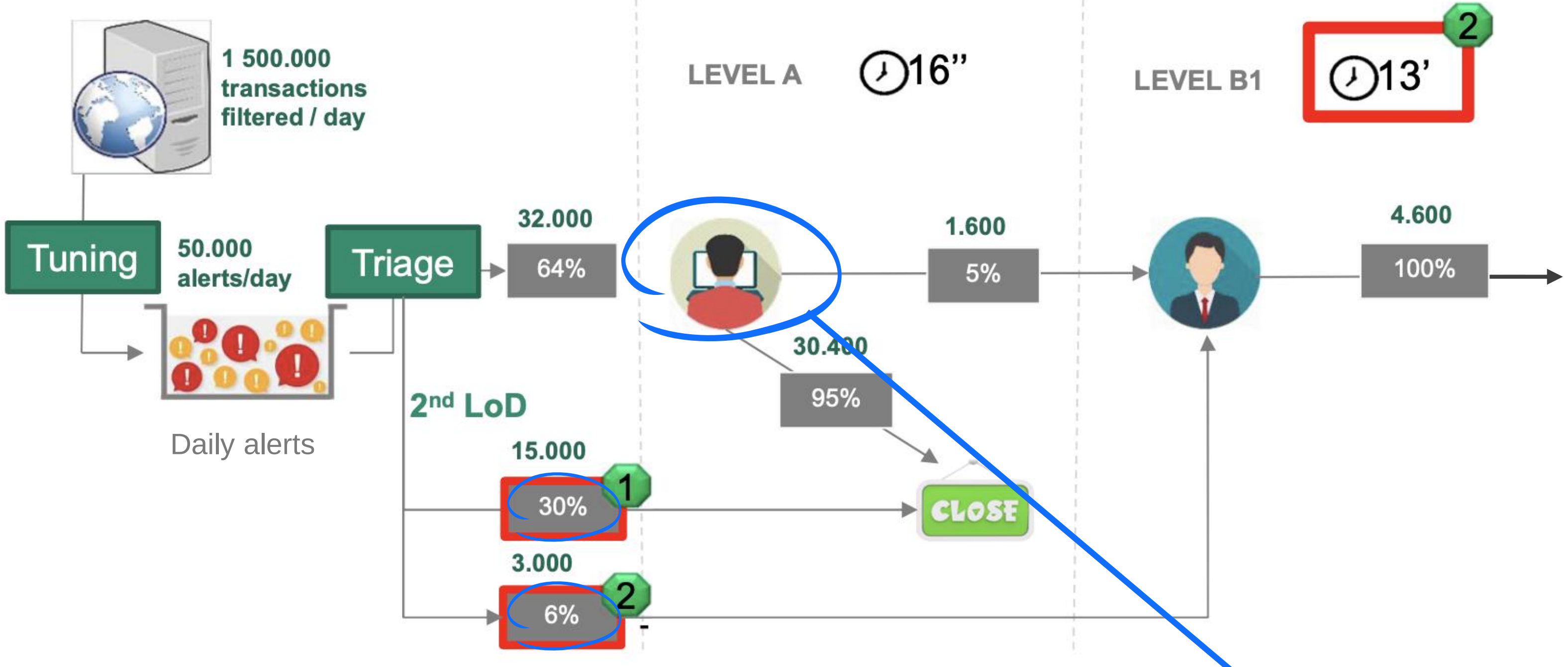
Reminder of Accuracy

Leading to "**Why are we using AI?**"

(1) IBM Global AI Adoption Index 2021

A customer case – Financial Sanctions

IBM team helps this client use ML to predict false positives, with satisfying algorithmic performances but...

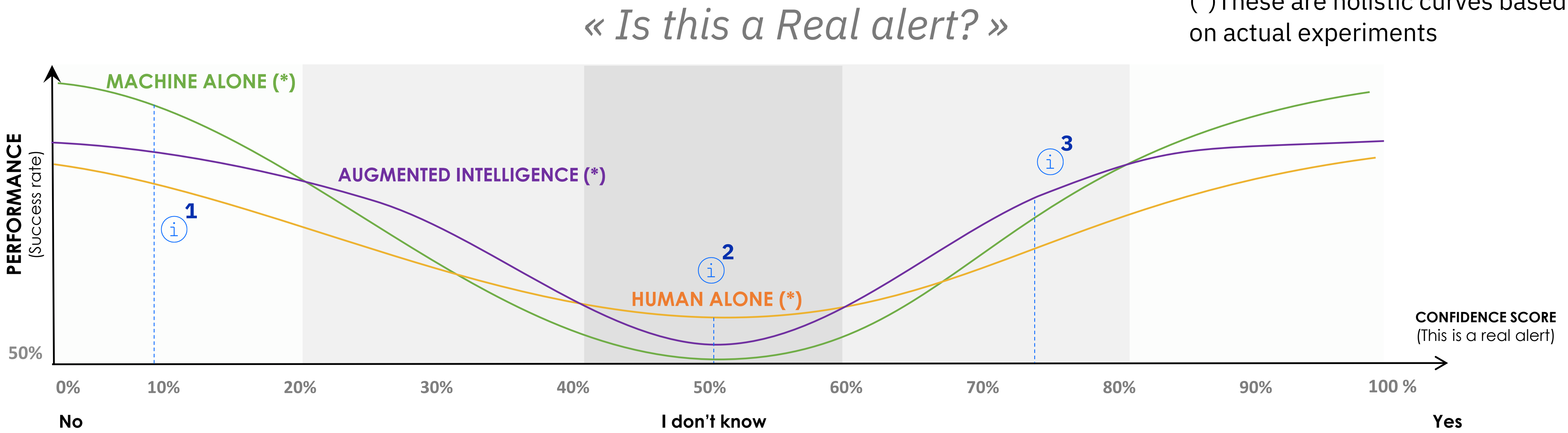


When should these decisions be automated ?

How should analysts leverage the AI recommendations ?

The **‘When’**

(*)These are holistic curves based on actual experiments



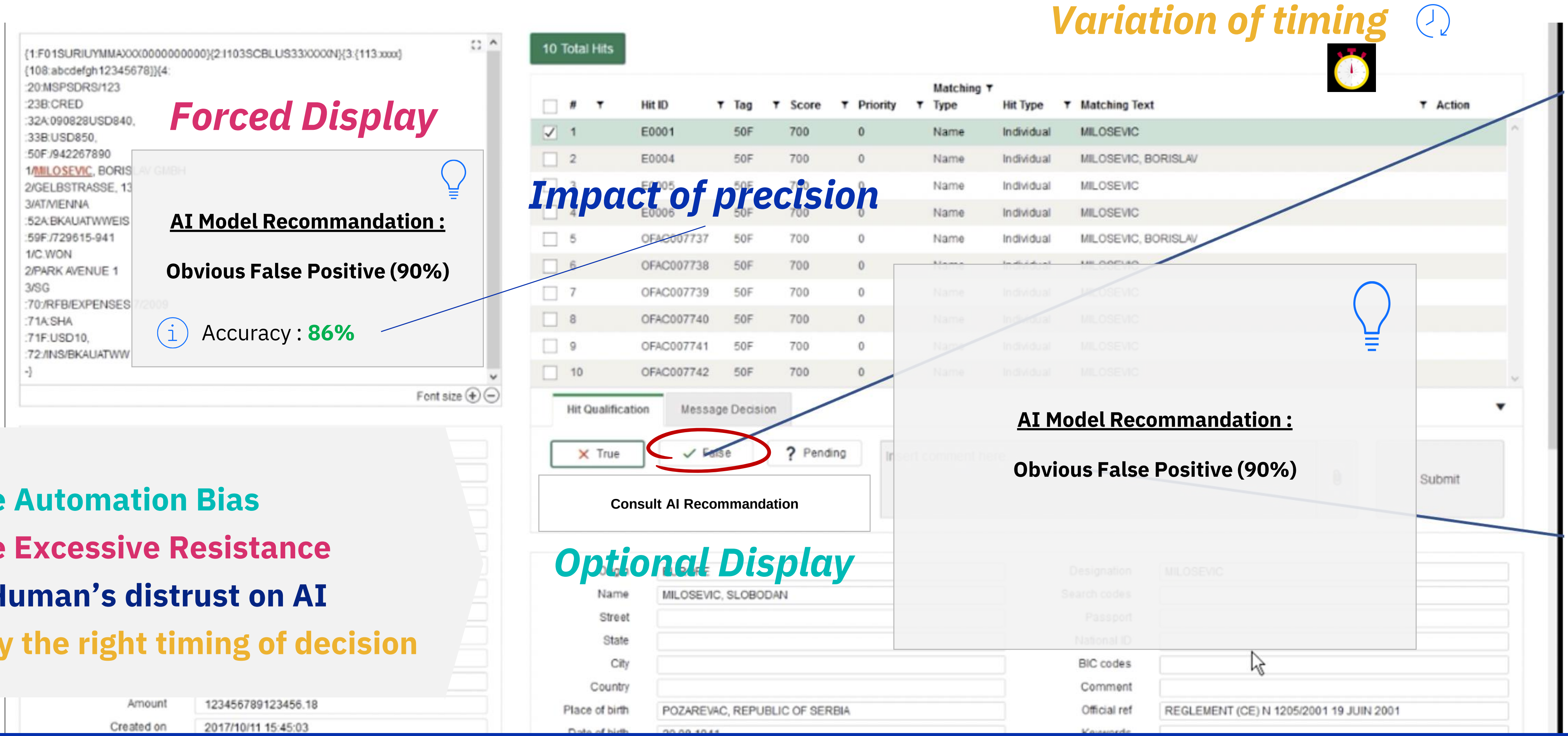
1 *AI is the best decision process*

2 *Human is the best decision process*

3 *The collaboration Human/Machine is the best decision process*

According to the level of confidence of the algorithm, we are able to define what is **the best decision process** to maximize the performance

The ‘How to’



- Reduce Automation Bias
- Reduce Excessive Resistance
- Raise Human’s distrust on AI
- Identify the right timing of decision

Knowing how each display impacts potential bias based on the algorithmic confidence level, we can define the **best UX for an optimal AI/Human collaboration with minimal bias**

IBM

The ‘Why’… with evidences, metrics and facts

Why		Because	
	Do we use AI in this process?		it is the **most efficient solution in X%** of the cases
	Do we use it here?		it outperforms Humans **at this level of confidence 90%**
	Should we trust it?		It helps human **take better decision in X%** of the cases
	Should it decide?		it is the **most efficient solution in X%** of the cases
	Is it not replacing a human?		**Humans take the best decision in X%** of the cases
	Is my decision influenced by AI recommendation		It helps **improve yours decisions by X%**
	Does my interface look like this ?		It reduces **Human Cognitive Bias by X%**
	Are humans central in this decision?		**It outperforms AI by X% in Y% of the cases**

Concretely, we replace **subjective impressions** and feedbacks with **Quantitative measures**

The client's **'Benefits'** and **'Value'**- Cross use cases & industries

Justify investments in AI with performance-based evidence

Providing facts and metrics, ObjectivAIze allows organizations to objectively assess the relevance of AI and the associated expected gains. Organizations can now take informed decisions when it comes to integrate AI in critical processes

Justify the use of AI towards regulatory bodies

Providing solid evidences of the relevance of AI in critical processes, Organizations can justify why they are using AI towards regulators, increasing their overall compliance and security.

Human resources are leveraged at their best, in full transparency

Knowing when Humans are optimal allows Organizations to delegate tedious tasks to AI and let collaborators focus on where they bring most added-value.

Conclusion

When to use A+H?

How to use it?

What is the performance gain?

Augmented Decision-making implies a sharing of responsibility between the system designer, implementer and the human in charge of the decision.

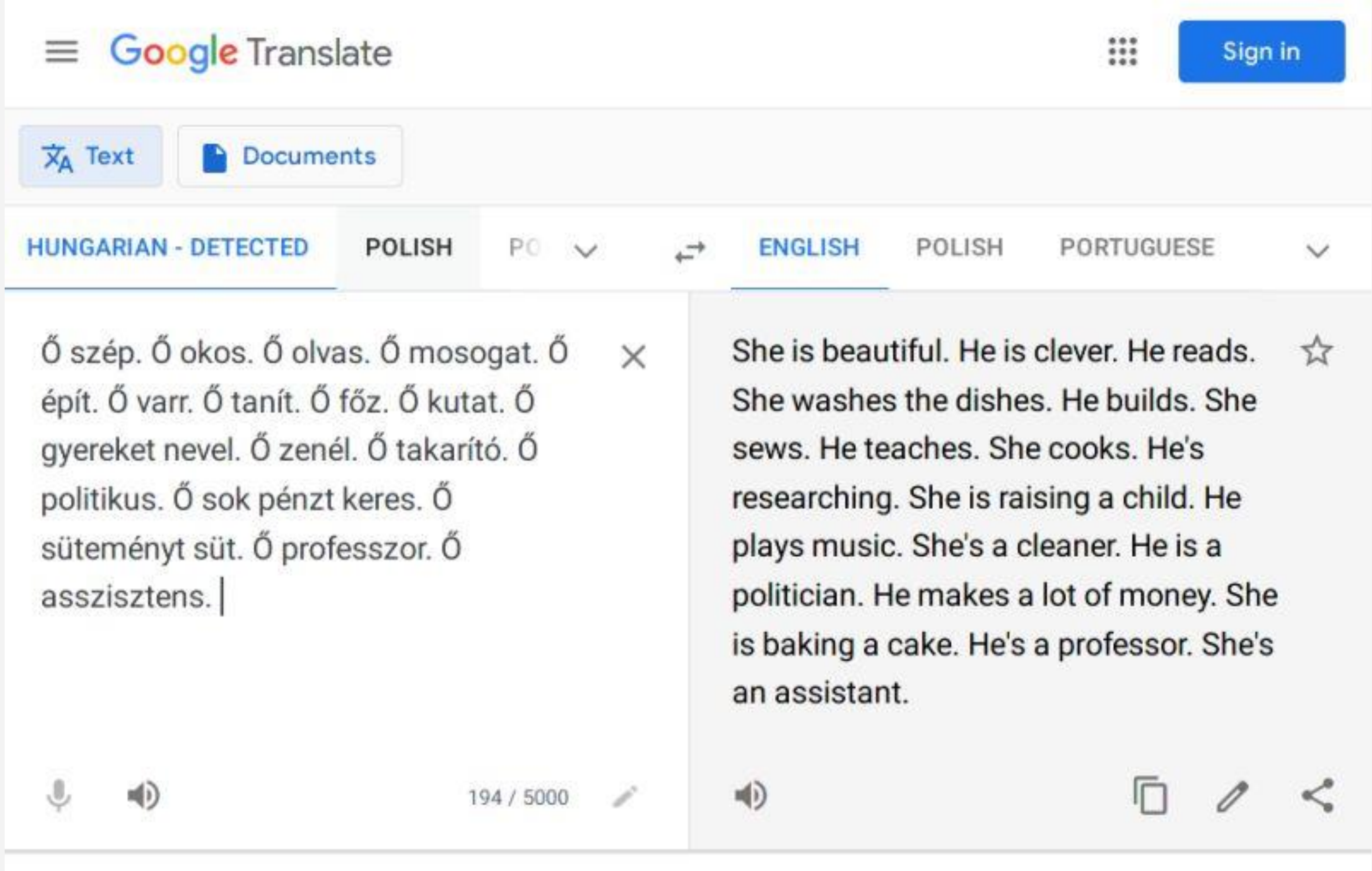
Specially if this is provided in our products as generic features.

-> like in avionics, we can envision a future sharing of liability between the engineer/designer and the user of decision support systems.

“guidelines”, “checklists”, “participatory design” won’t address this. We need engineering tools, metrics and methods to address those issues.

Towards an *objectivation* of AI Ethics.

Some interesting issues.



Algorithms can rub our collective hypocrisy to our face.

Some thougts on the use of Digital technologies to augment our understanding of Ethics

For a long time, it has been argued that discriminatory biases are common in Sensitive Decision Automation and Decision Support. This is for a large part the motivation of the EU proposal for a regulation of AI. The COMPAS case study made the news a while ago. This is a new case study, built with the same methodology as the propublica article on COMPAS, that highlights a systemic racism in granting mortgages, enforced instead of being corrected by algorithms trained on a dataset of past human decisions. https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms

What these stories reveal is not exactly that engineers designing the system failed. Like in the COMPAS case (which, btw, is still in use), the designers and the product owners have argued that their system only reflects the practices of the past, and that humans can still exercize their judgment (and probably are in the case of COMPAS).

Rather, to me, they raise 2 more interesting observations:

1- Algorithms have the power to show the discrepancy between our (collective) attitudes and our behaviors. What social psychology and behavioral economics have studied for a long time at the individual level can be shown at a collective level. **Algorithms can rub our collective hypocrisy to our face.**

2- Shouldn't sensitive automated decision-making be conducted by rule systems, as is done with our IBM Decision Automation products (shameless plug)? Those are a priori not subject to unconscious/non explicit biases, or by the contextualized repetition of past decisions. Is there a sweet spot to find between machine learning and decision logic to handle those sensitive types of decisions?

To me, this is a big prospect of Machine Learning, provided we use it for the right purpose, to reveal our collective biases rather than to amplify them; leveraging tools such as IBM AI 360 Fairness toolkit and not blindly; or using it in association with decision logic that guards against hidden but systemic deviations from our ethical values. See also

