IN-THE-LOOP / ON-THE-LOOP: **COMMENT CHOISIR?**

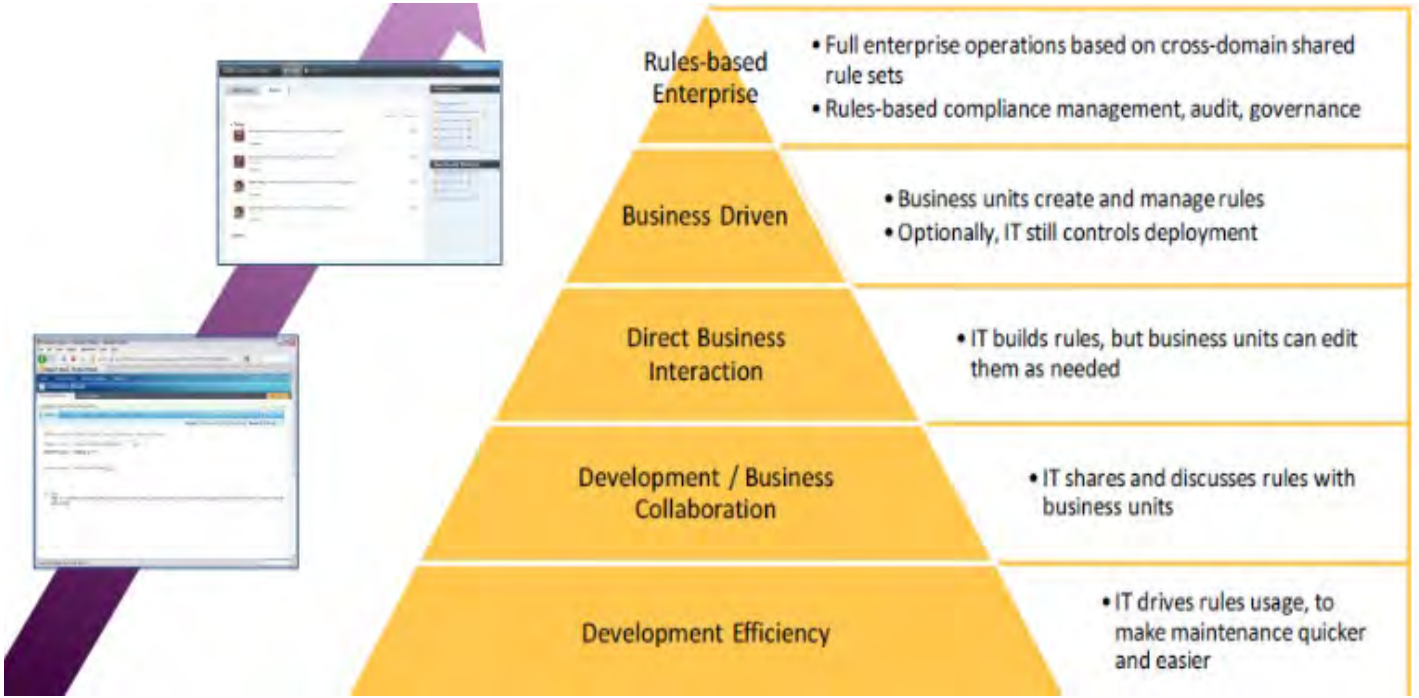
Objectiver l'usage de la décision automatisée et de l'aide à la décision dans les processus métier tels que le filtrage d'alertes AML.

Thomas Baudel, CTO AI Decision Coordination, IBM
Webinaire "Les Lundis de la Finance", Telecom ParisTech & ACPR, 6/03/2023

IBM

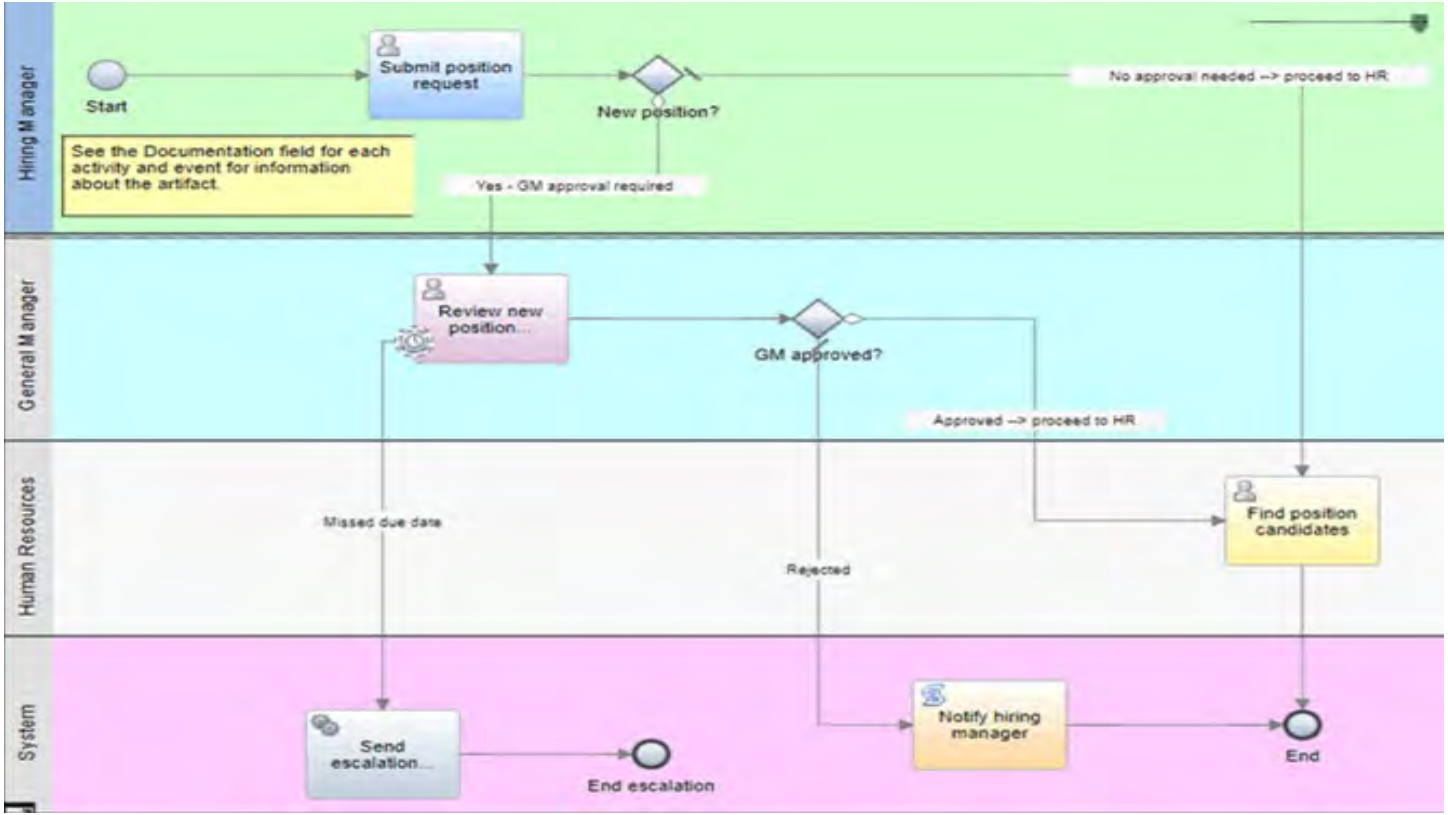
INGENIERIE DE LA DÉCISION

Programmer une entreprise comme on programmerait un robot



J. Taylor and N. Raden, Smart Enough Systems. Prentice Hall, 2004.

LA GESTION DE PROCESSUS MÉTIER ORCHESTRE LES FLUX D'INFORMATION DANS L'ENTREPRISE: TÂCHES ET DÉCISIONS



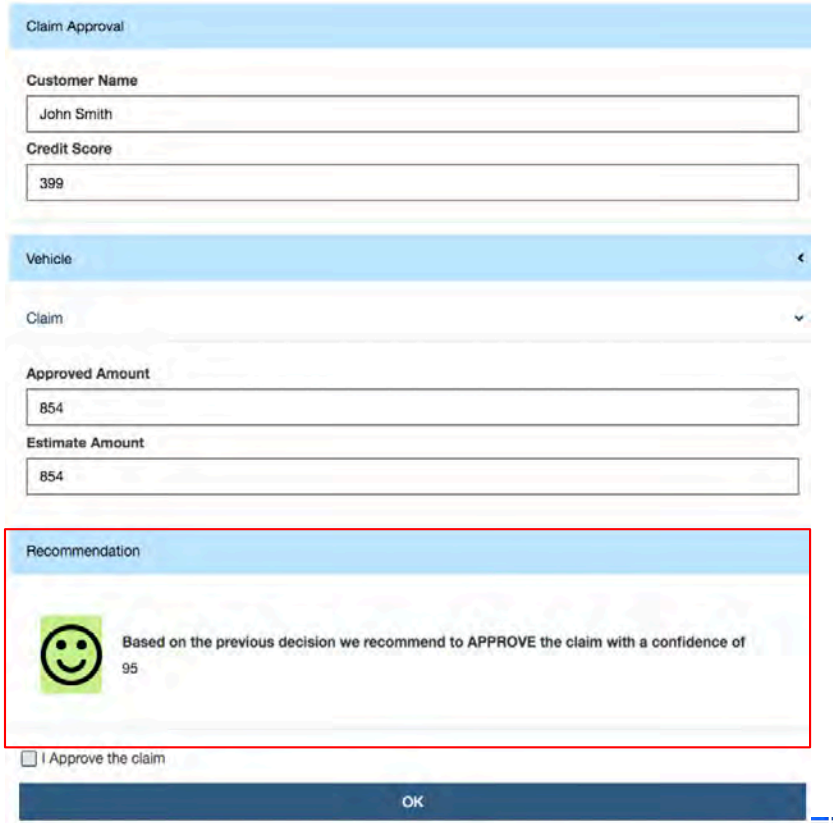
AIDA: ARTIFICIAL INTELLIGENCE FOR DECISION AUTOMATION

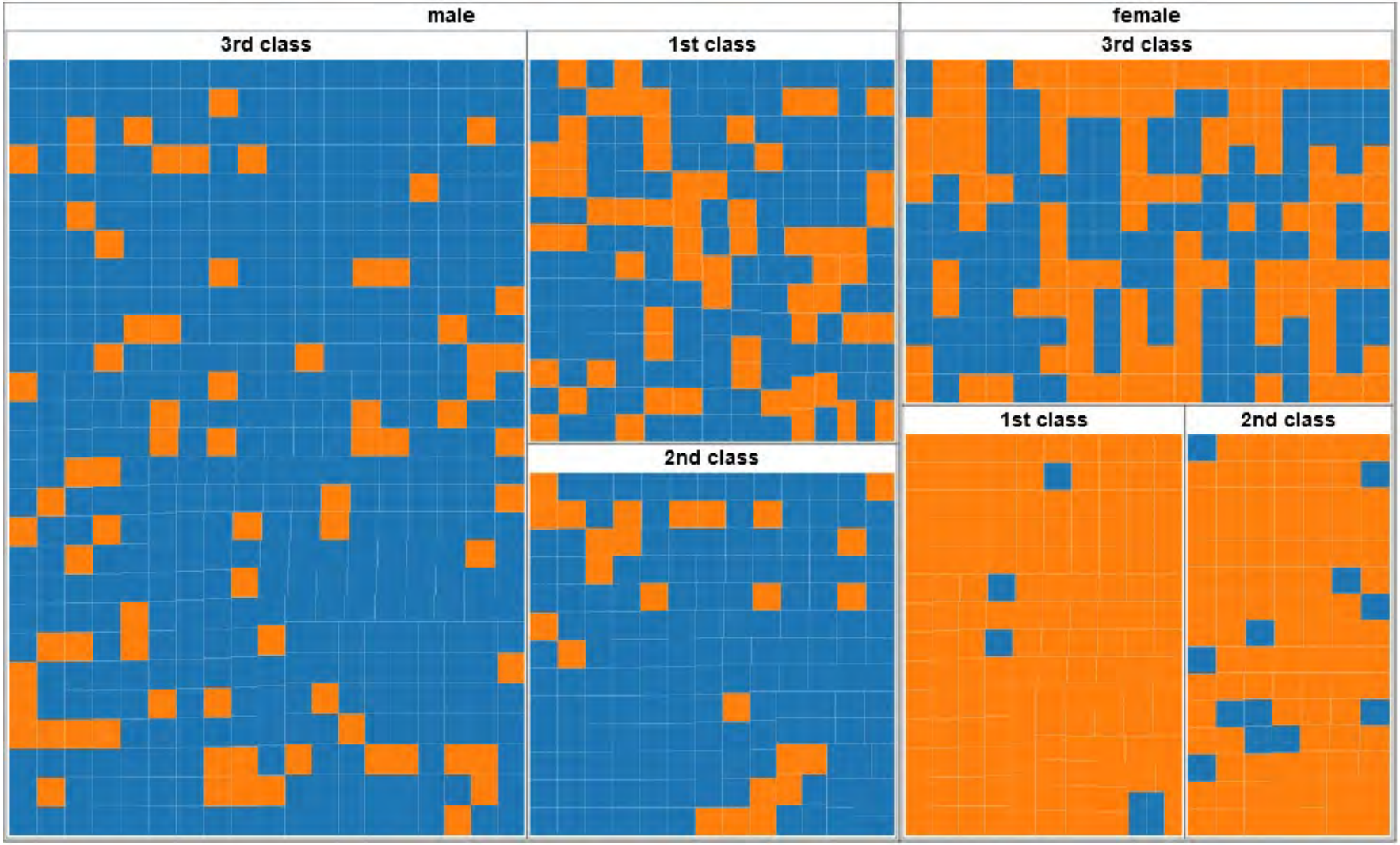
On capture les flux de décisions: apprenons des décisions passées pour proposer une recommandation?

Exemple: validation de demandes de remboursement de frais médicaux.

En principe, une application très simple de l'apprentissage automatique, applicable de façon générique.

Des questions d'ingénierie se heurtent à un obstacle non technologique: qui est responsable des décisions prises?




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

Passenger data:

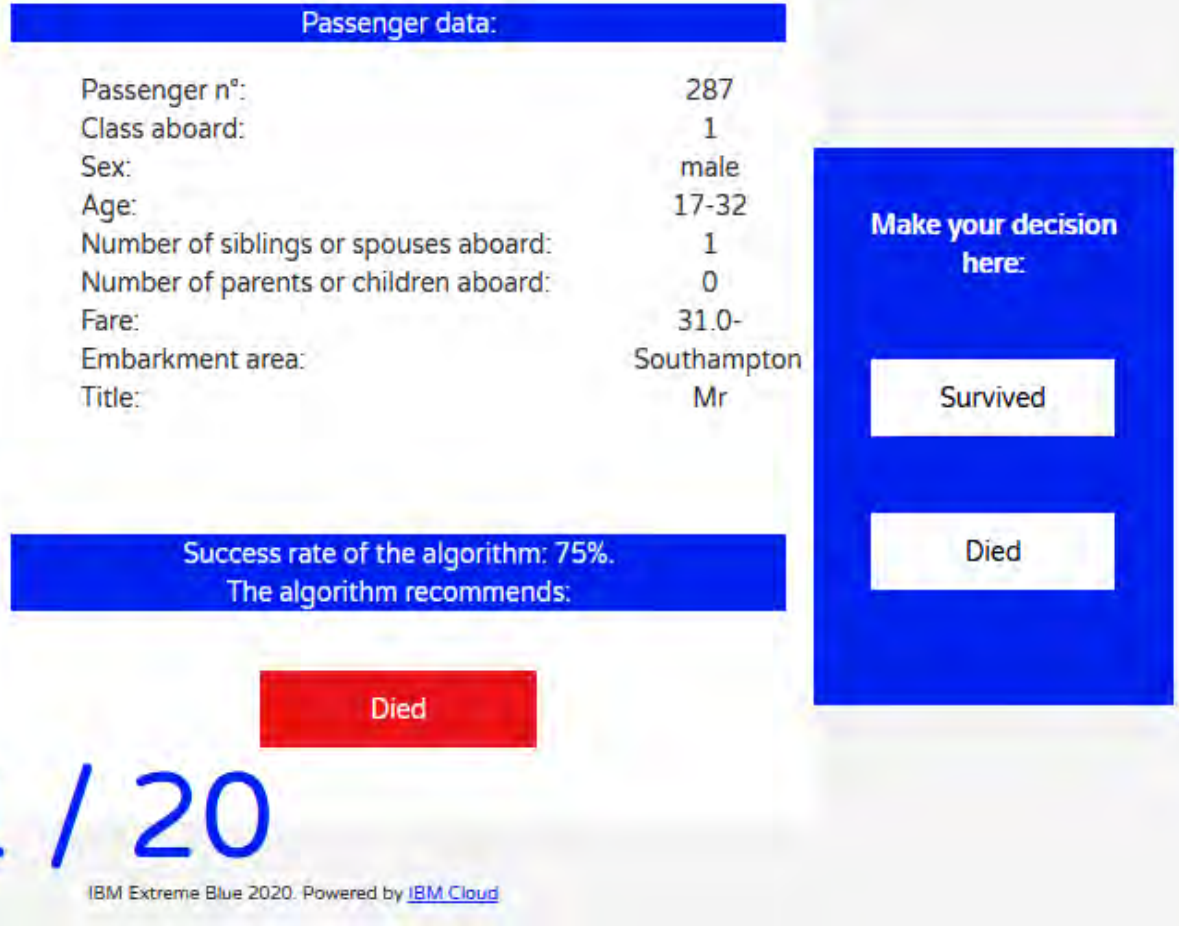
Passenger n°:	179
Class aboard:	1
Sex:	female
Age:	49-64
Number of siblings or spouses aboard:	0
Number of parents or children aboard:	2
Fare:	31.0-
Embarkment area:	Cherbourg
Title:	Mrs

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
Make your decision here:
Survived
Died
Success rate of the algorithm: 75%.
The algorithm recommends:
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

Make your decision here:

Survived

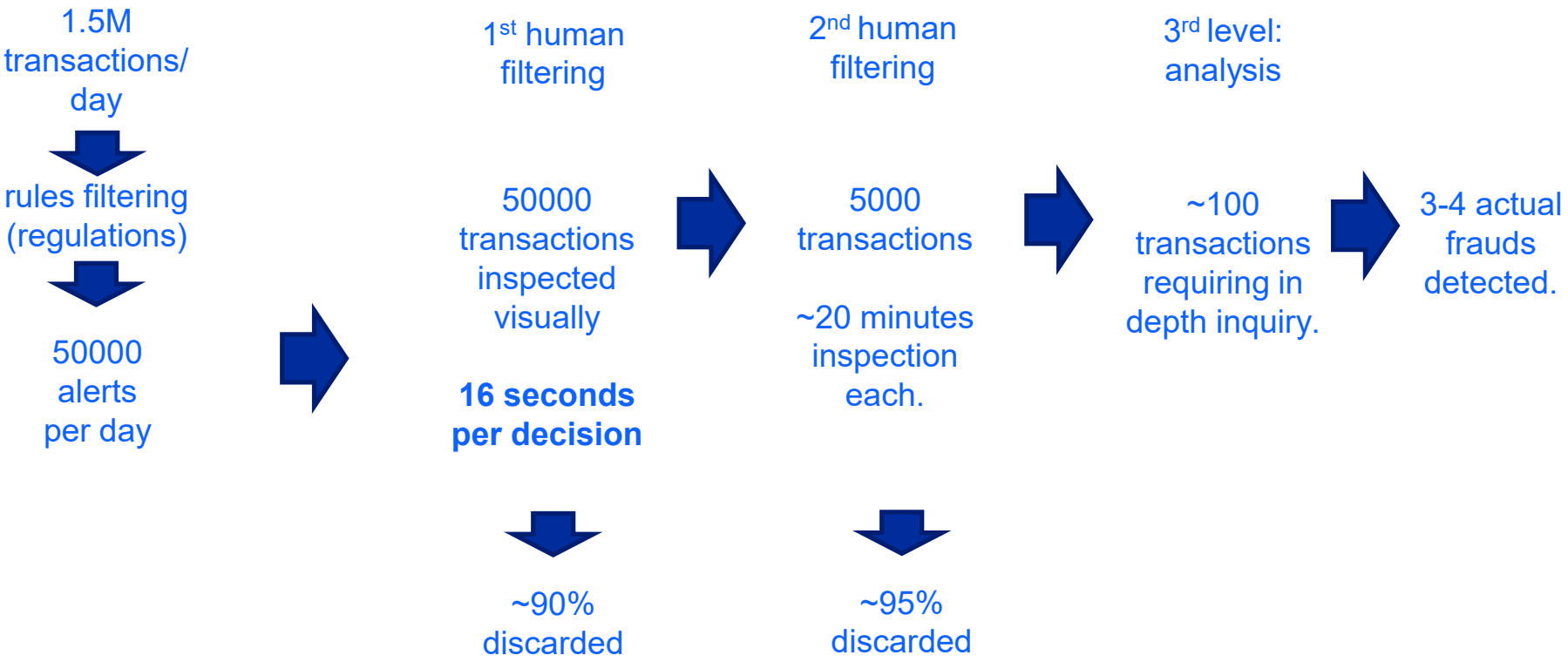
Died

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

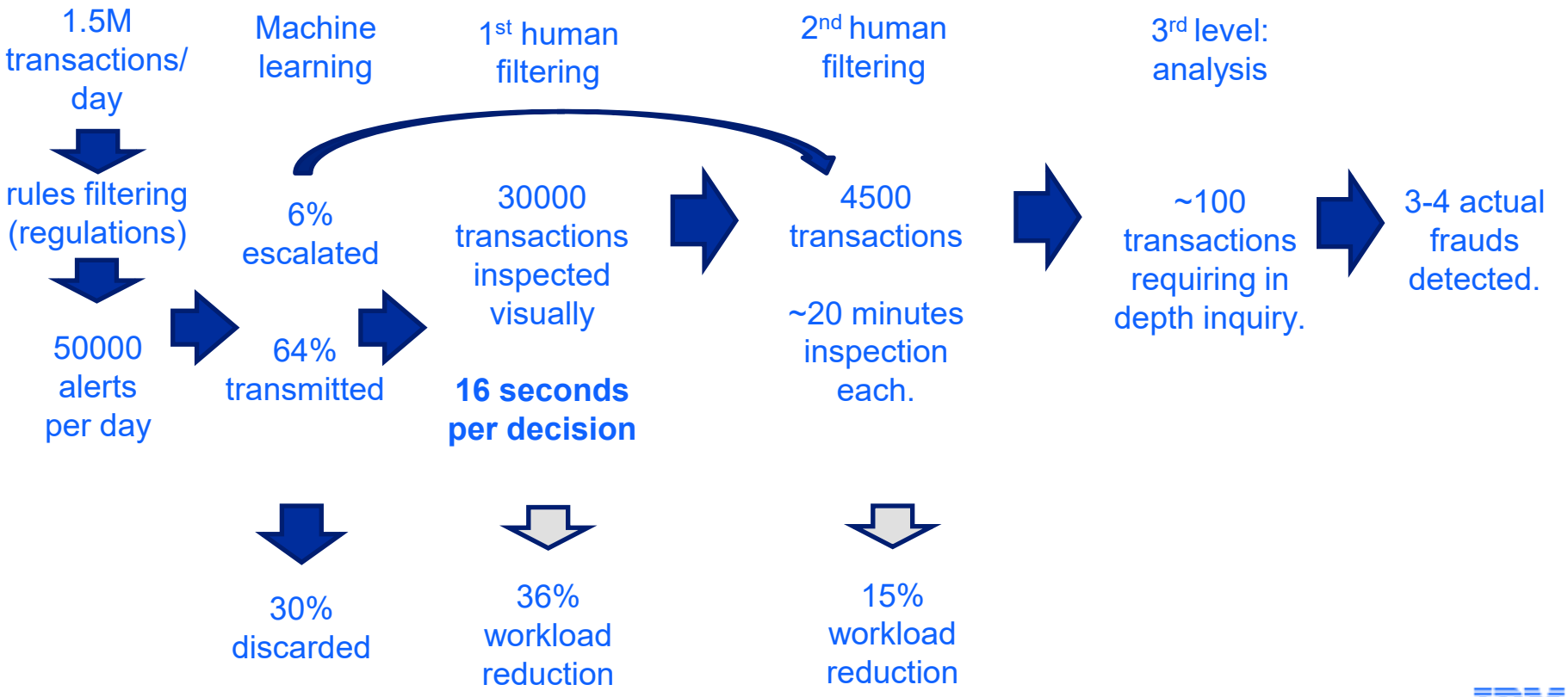
Survived

4 / 20

CAS D'USAGE RÉEL: FILTRAGE D'ALERTES SUR TRANSACTIONS FINANCIÈRES (AML/SFI/FRAUDE…)

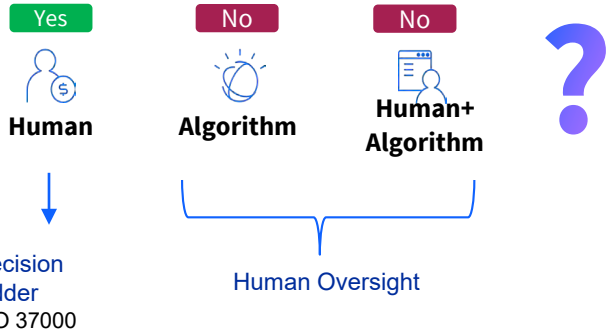


CAS D'USAGE RÉEL: FILTRAGE D'ALERTES SUR TRANSACTIONS FINANCIÈRES (AML/SFI/FRAUDE…) (2)



Contrôle Humain/Supervision: un besoin de l'industrie

QUI décide?



COMMENT l'algorithme influence la décision?

- **Preservation of Autonomy**
- **Automation bias**
- **Order or similarity bias**
- **Decision Fatigue**
- **Timing effect**
- **Expertise effect**

Human Agency

QUI est responsable de la décision?

Ont-ils les moyens d'assumer cette responsabilité?

85% des projets de déploiement échouent.

https://research.aimultiple.com/ai-fail/

Contrôle Humain/Supervision: une exigence réglementaire

Regulations on Decision Automation per industry sector

Existing, long-standing, per-Industry regulations: **Safety regulations** in plenty of industries make provisions for automated decision governance: 70 years of industry practice (1951) ; **Finance & insurance** industry also have in-place regulations for automated decision: ~35 years of industry practice (1987).

AI regulation projects

Canada (AIDA): Oversight mostly seen in the context of risk management + ministry of industry oversight
UK: “Define legal persons' responsibility for AI governance”, “Clarify routes to redress or contestability”, sector-specific approach.

US, China…

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. 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] […]

CEN/CENELEC: standardization request from EU (JTC 21, WG3, AHG7), article 2.5: Human Oversight.

This (these) European standard(s) or European standardisation deliverable(s) shall specify measures and procedures for human oversight of AI systems which are:

(a) identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service;

(b) identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user.

These shall include measures enabling users to understand, monitor, interpret, assess and intervene in relevant aspects of the operation of the high-risk AI system.

This (these) European standard(s) or European standardisation deliverable(s) shall also define, where justified, appropriate oversight measures which are specific to certain AI systems in consideration of their intended purpose. With respect to AI systems intended for remote biometric identification of persons, human oversight measures shall inter alia ensure that no action or decision is taken by the user on the basis of the identification resulting from the system

Contrôle humain/supervision: *Qui* est en charge?

i.e. aurait pu prendre une décision informée et non contrainte aboutissant à un résultat différent?

Dans les organisations humaines

Oversight is one of the 'principles of governance' covered in depth in ISO 37000:2021, 6.4.

-> Management stays on top of the organization

ISO/IEC 38507:2022 Information technology — Governance of IT

Oversight
monitoring of the implementation of *organizational and governance policies* and management of associated tasks, services and products set by the organization, in order to adapt to changes in internal or external circumstances
Note 1 to entry: Effective oversight needs general understanding of a situation. -> ISO 37000

-> Management stays on top of IT.

Pour le législateur: respect de l'autonomie humaine

EU HLEG AI, 2018, 1st guideline:
Human Agency & Oversight
AI systems should support human autonomy and decision-making, as prescribed by the principle of **respect for human autonomy**. […]

-> Humans as a whole stay in charge.

Human-in-the-loop (HITL) (decision support): capability for human intervention in every decision cycle of the system, ***which in many cases is neither possible nor desirable.***

Human-on-the-loop (HOTL) (decision automation)

Human-in-command (HIC)

-> Various modalities of what "stay in charge"means.

Pour le technologue: niveau d'autonomie de la machine

ISO/IEC 22989 – concepts : *focuses on the properties of the machine*, not on human autonomy.

Autonomy/autonomous
characteristic of a system that is capable of modifying its intended domain of use or goal without external intervention, control or oversight

Machine autonomy, table 1:
NOTE In jurisprudence, autonomy refers to the capacity for self-governance. In this sense, also, "autonomous" is a misnomer as applied to automated AI systems, because even the most advanced AI systems are not self-governing. Rather, AI systems operate based on algorithms and otherwise obey the commands of operators. For these reasons, this document does not use the popular term autonomous to describe automation

ISO TS8200 - Controllability: focuses on technical properties of the AI and its context of use, not on *who/how/at what level* control is exercised.

Dans les sciences humaines et sociales

multiple sources*:* [human] autonomy is the capacity to make an **informed, uncoerced decision.**

-> *informed*, leads to explainability, transparency and related concepts

-> *uncoerced*, leads to:

- Free of making choices (hierarchy)
- Uninfluenced (cognitive biases, situational constraints)

In a professional context, when a decision has been made, the issue becomes:

Who was in a position to make an informed, uncoerced decision that would have made a difference?

Le contrôle humain s'exerce sur une hiérarchie de niveaux: contrôle direct, supervision, impacts métier, impacts sociétaux…

THÉORIE DE L'ALLOCATION DE FONCTIONS (HABA – MABA, 1951)

Humans appear to surpass present-day machines in respect to the following:

1. *Ability to detect a small amount of visual or acoustic energy*
2. *Ability to perceive patterns of light or sound*
3. Ability to improvise and use flexible procedures
4. *Ability to store very large amounts of information for long periods and to recall relevant facts at the appropriate time*
5. *Ability to reason inductively*
6. Ability to exercise judgment

Present-day machines appear to surpass humans in respect to the following:

1. Ability to respond quickly to control signals and to apply great force smoothly and precisely
2. Ability to perform repetitive, routine tasks
3. Ability to store information briefly and then to erase it completely
4. Ability to reason deductively, including computational ability
5. Ability to handle highly complex operations, i.e. to do many different things at once.

Fitts PM (ed) (1951) Human engineering for an effective air navigation and traffic control system. National Research Council, Washington, DC

HABA – MABA REVISITÉ

Humains et Algorithmes se complètent:

- Les humains exploitent un contexte, cherchent de l'information externe, adaptent les critères de décision.
- Les algorithmes exploitent plus d'information, calculent plus vite à un coût inférieur.

Et ont des faiblesses différentes:

- Les algorithmes gèrent mal les irrégularités, les exceptions, la nouveauté.
- Les humains sont irréguliers dans leur performance: fatigue, biais cognitifs, parti-pris…

IBM

THEORIE DE L'ALLOCATION DE FONCTIONS POUR L'AIDE À LA DÉCISION: LORSQUE L'HUMAIN EST DE TROP…

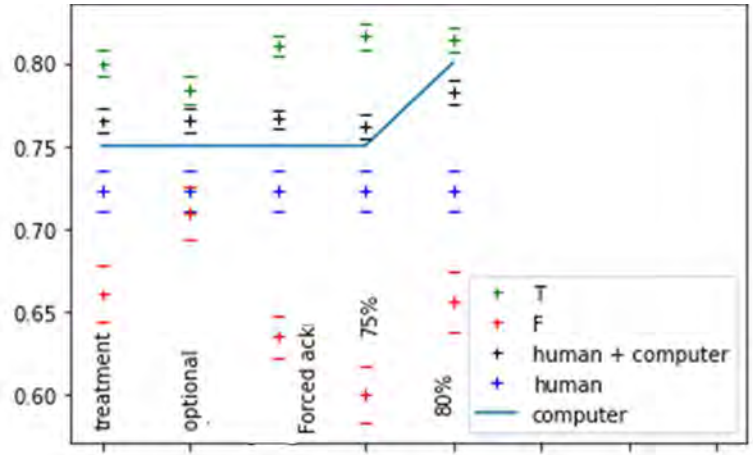
La complémentarité est possible, dans certaines limites:

[Onnash]: if the algorithm performance is < 70%, it is counter-productive as a recommender.

[Starke & Baber, Green et al.]: when performance of the algorithm is > 80%, then performance is *degraded* by human intervention (in select tasks): involving a human decision-maker is *counter-productive.*

[Baudel et al.]:

- Between 70% and 80%, there is a (small) window where collaboration is possible.
- Automation bias is reduced when recommendation is available only on demand.

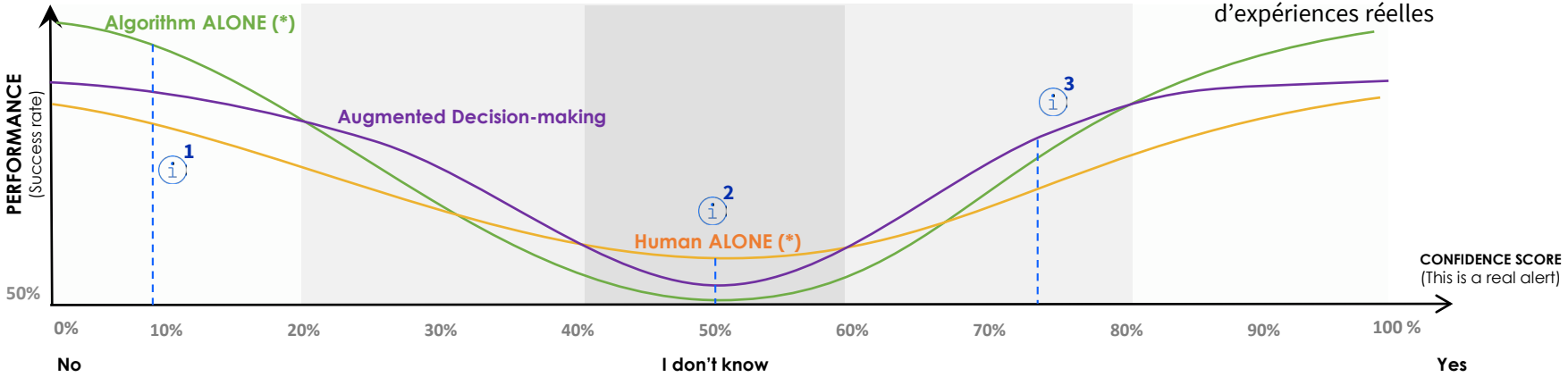


Linda Onnasch, Crossing the boundaries of automation—Function allocation and reliability, International Journal of Human-Computer Studies, Volume 76, 2015, Pages 12-21, ISSN 1071-5819, https://doi.org/10.1016/j.ijhcs.2014.12.004.
Sandra Dorothee Starke, Chris Baber, The effect of known decision support reliability on outcome quality and visual information foraging in joint decision making, Applied Ergonomics, Volume 86, 2020, 103102, ISSN 0003-6870, https://doi.org/10.1016/j.apergo.2020.103102.
Baudel T., Verbockhaven M., Cousergue V., Roy G., Laarach R. (2021) : Measuring Performance and Biases in Augmented Business Decision Systems. In: Ardito C. et al. (eds) Human-Computer Interaction – INTERACT 2021. INTERACT 2021. Lecture Notes in Computer Science, vol 12934. Springer, Cham. https://doi.org/10.1007/978-3-030-85613-7_22

PERFORMANCE EN FONCTION DE L'INDICE DE CONFIANCE (ALGORITHMIQUE)

« Est-ce un vrai positif ?»

(*) courbes idéalisées, issues d'expériences réelles



1 *Algorithm 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

APPORTS DE LA MÉTHODE

%

Justification des investissements

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.

Justification aux instances de régulation

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.

L’humain valorisé

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

IBM

CONCLUSION

Aide à la décision et automatisation de la décision imposent un partage de responsabilités entre concepteurs du système, responsables métier, agents en charge des décisions individuelles, régulateurs et public dans son ensemble.

Pour déterminer le rôle de chacun dans la prise de décision, nous proposons des **métriques et une méthodologie objectivable** issue de la pratique des facteurs humains en sécurité industrielle.

Vers une *objectivation* de la mise en œuvre de l'IA.

Pour toute information supplémentaire: AIDC.Contact@ibm.com

IBM

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

IBM

HOW IS THIS ADDRESSED FOR NOW?

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

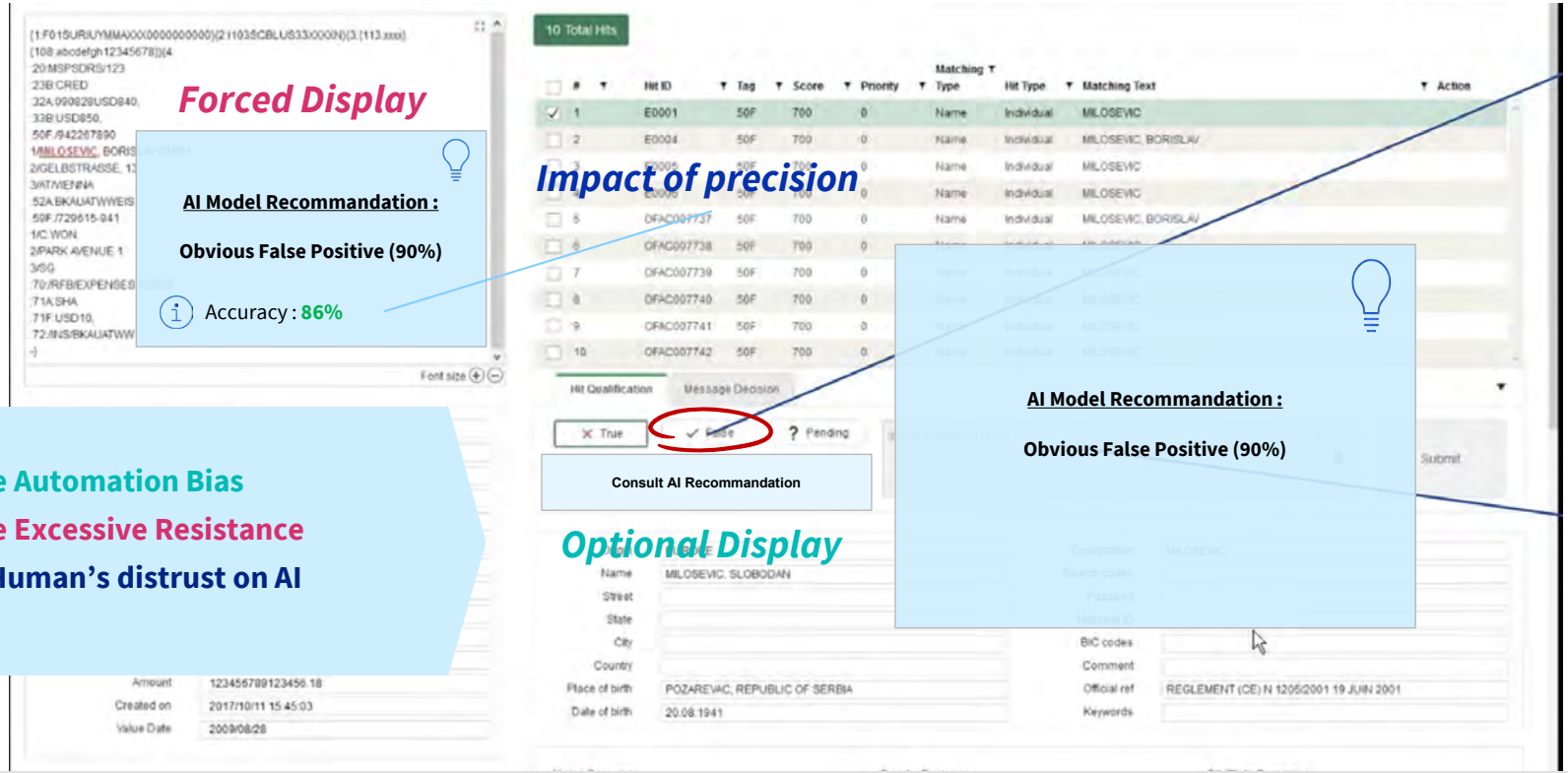
>> Perceptual effects

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

Work process changes

>> Risk replaced by uncertainty: acceptability issues

THE “HOW”



- **Reduce Automation Bias**
- **Reduce Excessive Resistance**
- **Raise Human’s distrust on AI**

Empirical experiments (A/B testing) allow understanding the impact of presentation on collaboration effectiveness