



Ressources rares, cerveaux fatigués, société vieillissante : une IA post-croissance ?

Pr. Patrick Le Callet

Keynote

10 juin 2025



institut
universitaire
de France



Nantes
Université



Human-Machine Experience & Cognitive MultiMedia Computing

Interdisciplinary team for fundamental research ...

Computer-Human
interface

Coding

X-realities



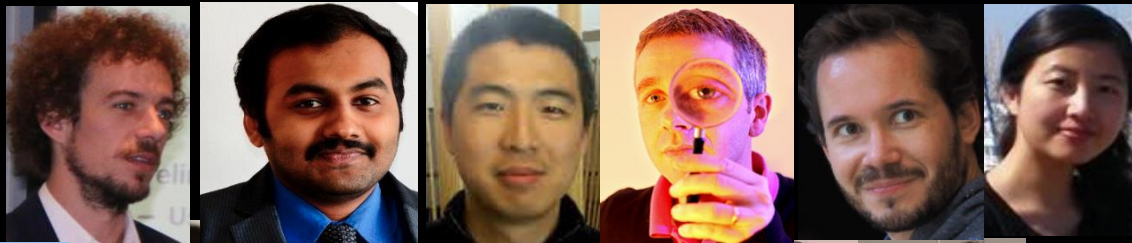
(IPI: Image
Perception &
Interaction)



Ophtalmology
Neurosciences

Design
Psychology

AI for
Human signals



...combined with innovation
facilities





société vieillissante

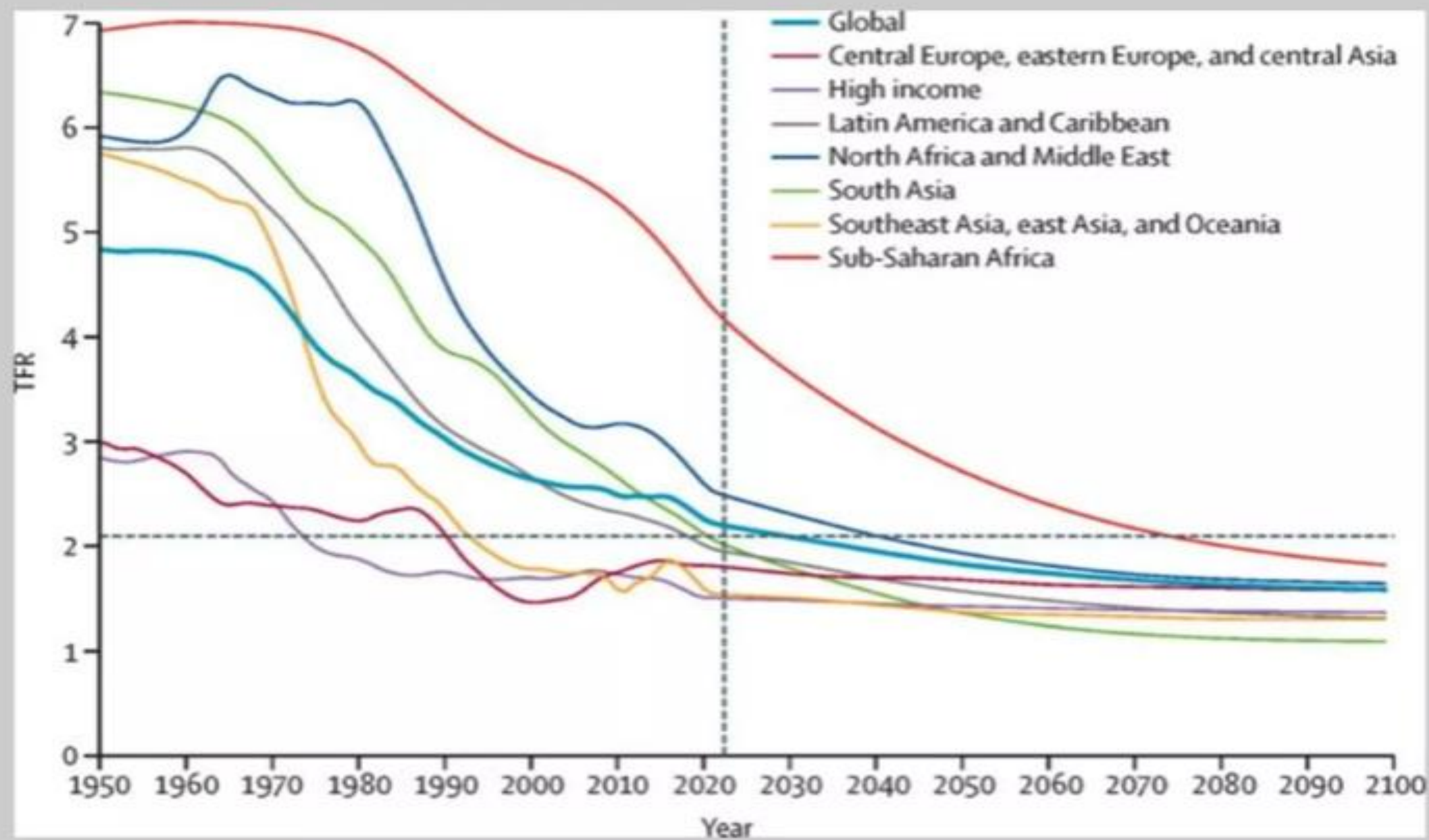


institut
universitaire
de France



Nantes
Université





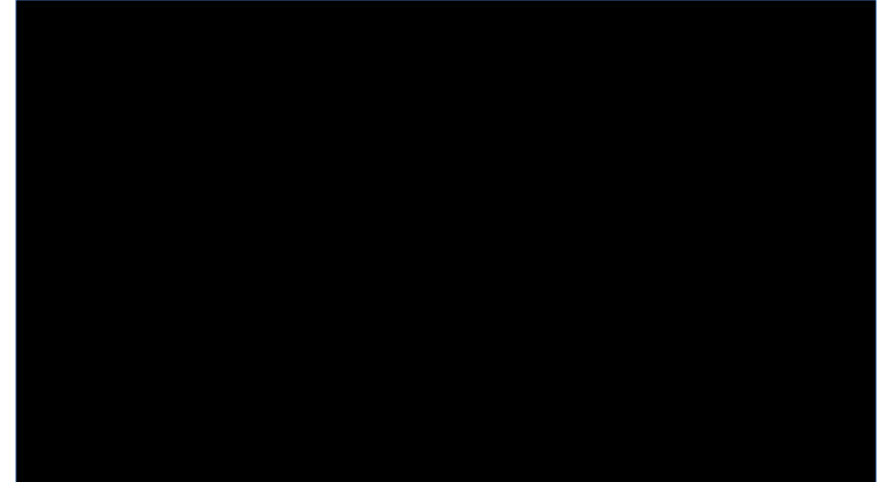
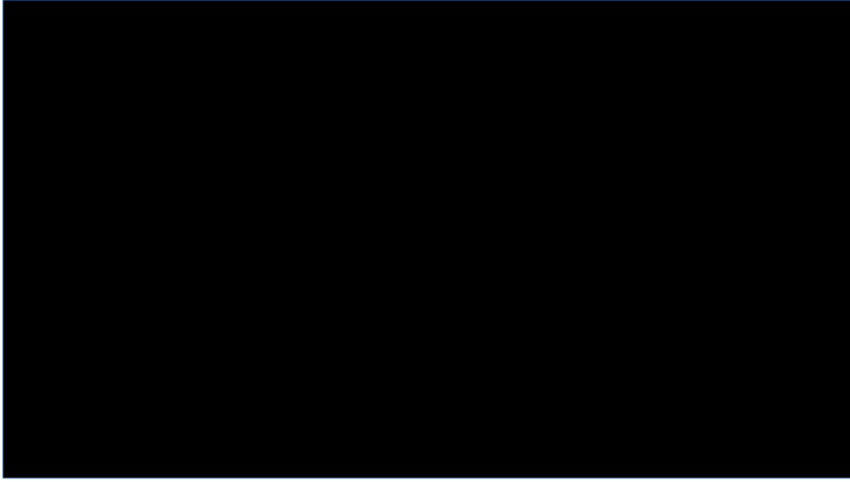
Evolution des taux de natalité 1950 - 2100 (projection) © The Lancet

Avènement de l'IA incarnée et cobotique accessible



robots programmés par des
ingénieurs

Apprentissage par imitation?



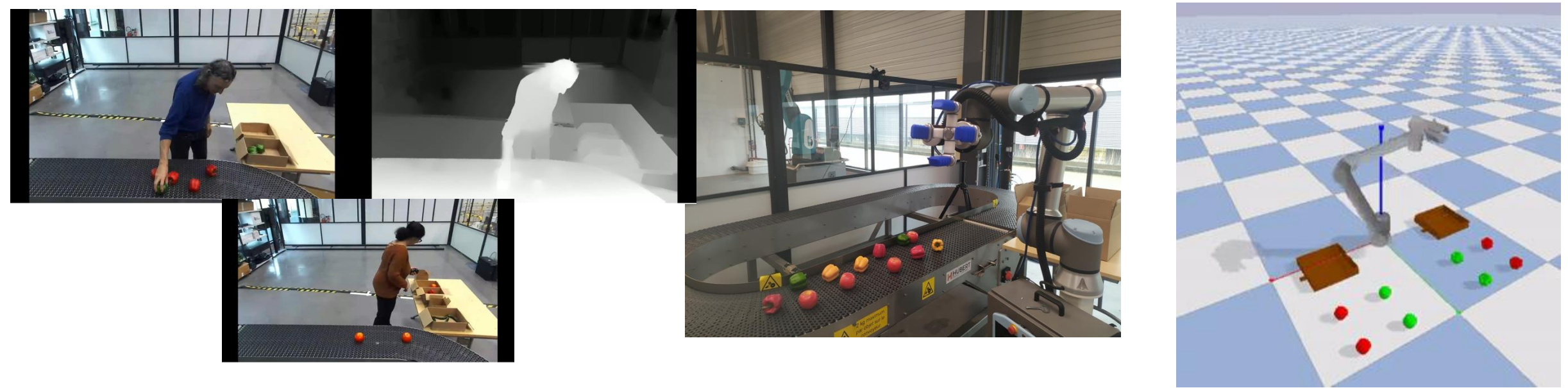
Avec les progrès de la vision par ordinateur



[1] Kormushev, Petar, Sylvain Calinon, and Darwin G. Caldwell. "Imitation learning of positional and force skills demonstrated via kinesthetic teaching and haptic input." *Advanced Robotics*

[2] Fu, Zipeng, Tony Z. Zhao, and Chelsea Finn. "Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation." *CoRL* 2024

Apprentissage par imitation (Imitation learning)



trajectoires « experts »

Extraction d'intention

Espace de
Représentation
IHO
(Interaction
Humain Object)

demonstration

Information visuelle

Compréhension de
Scènes

Apprentissage de tâches

Imitation learning

« experts » trajectories

demonstration

Visual Information

Intent extraction

Scene understanding

Representation
space of
HOI (Human
Object
Interaction)

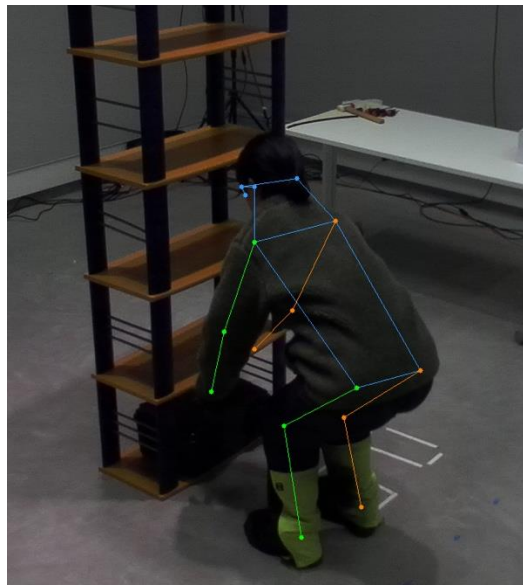
Yassine El Manyari:
Deep Reinforcement Learning
Transformer seq2seq
...

Task learning



2D pose estimation under strong occlusions

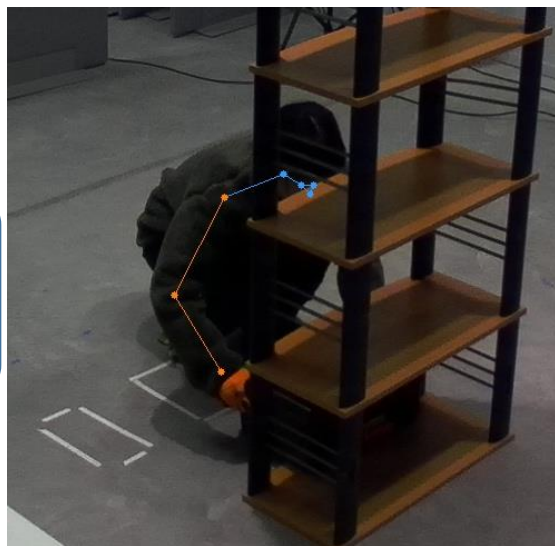
no
Occlusion



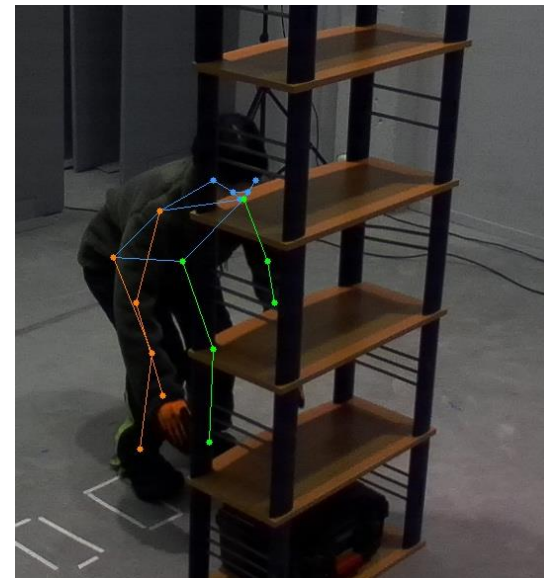
Occlusion
fail

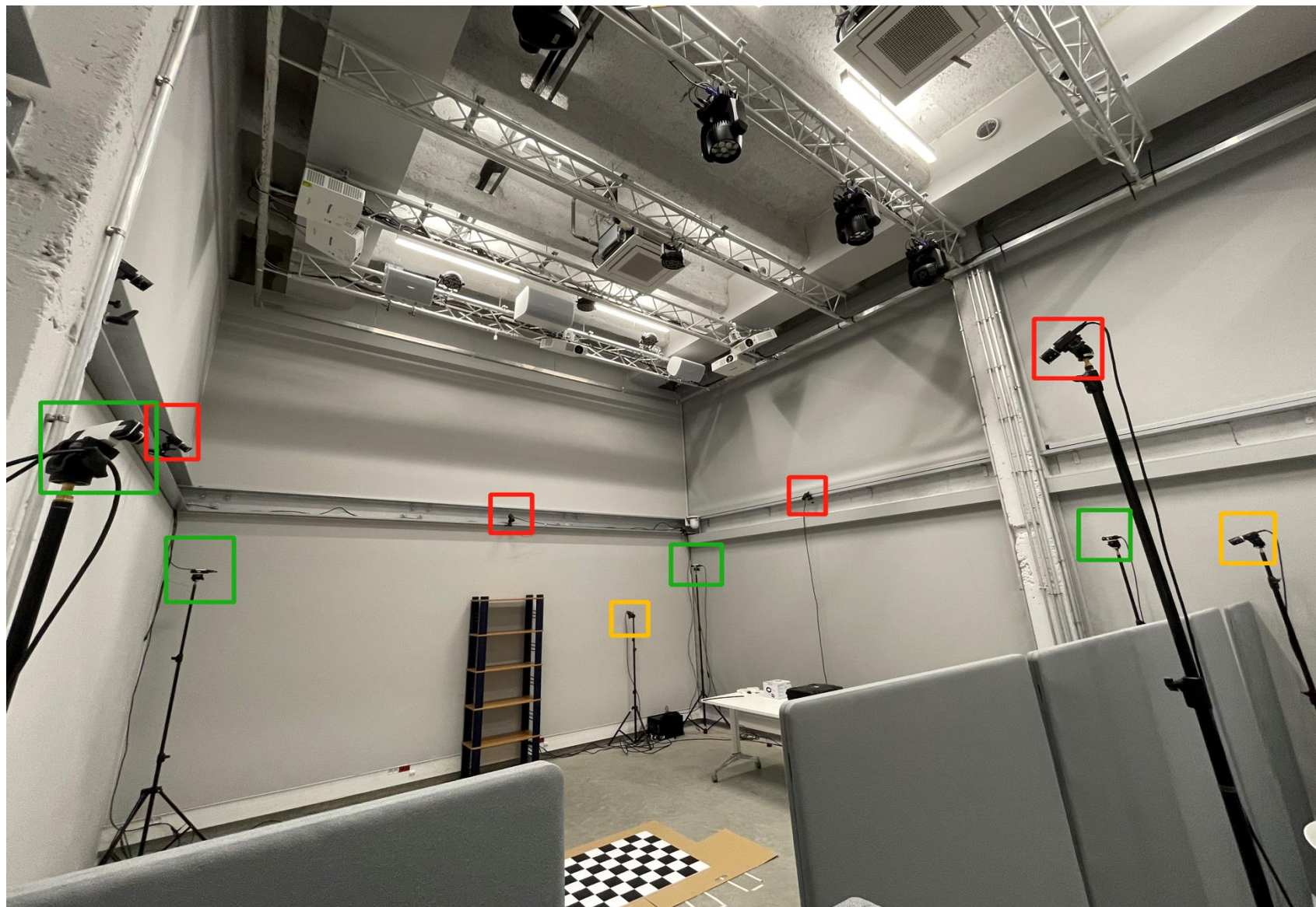


Occlusion
partial
success



Occlusion
success





- 8 Flir (RGB industrial cameras) with short focal length ;
For global view
- 4 zed (3D stereo-cameras) For global view
- 2 additional Flir with long focal length ;
For local view

[Evaluating 3d human pose estimation in occluded multi-sensor scenarios: dataset and annotation approach](#) K Riou, K Dong, Y Huang, K Subrin, P Le Callet, Y Sun, ICIP 2024

Computational Ergonomics

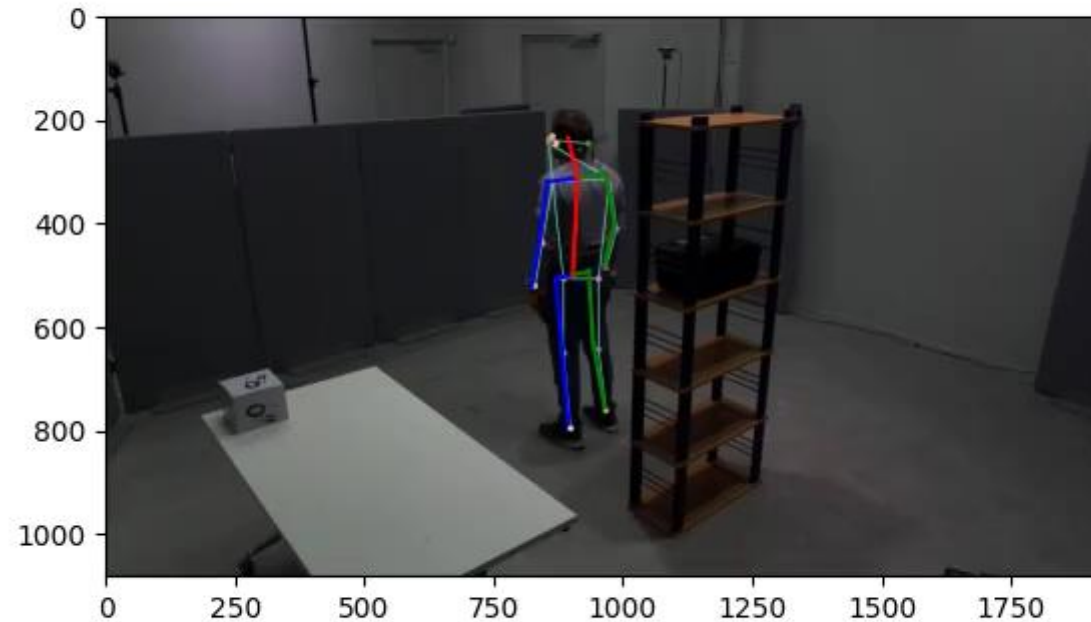


Screwing&Hammering examples, 6 « ergonomic » variations

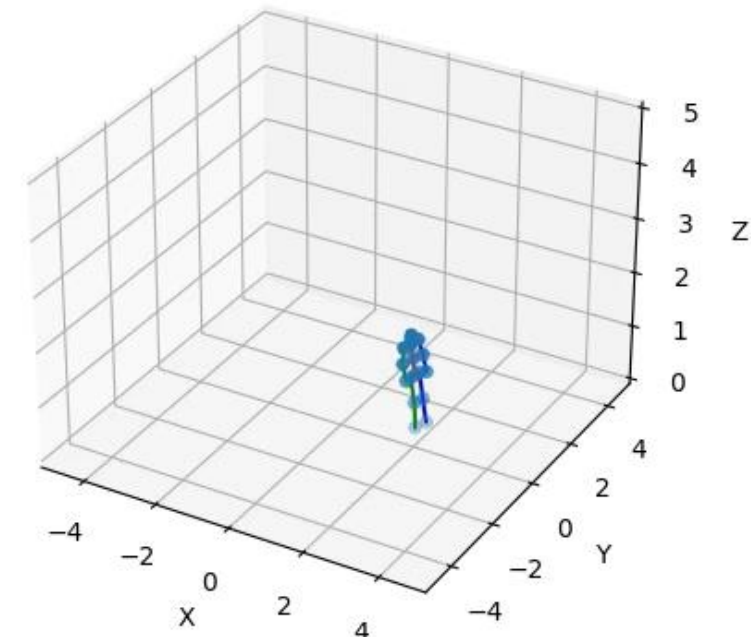


Triangulation in « highly » multi-view setup, as 3D human pose ground truth

3D pose Re-projected to 2D (thick colored lines) vs initial 2D (thin white lines)



Visualization for index 160



[Spatial-Temporal-Geometric Graph Convolutional Network for 3-D Human Pose Estimation from Multiview Video](#) K Dong, Y Zhou, K Riou, X Yun, Y Sun, K Subrin, P Le Callet

IEEE Transactions on Instrumentation and Measurement, 2024

[Geometric Consistency-Guaranteed Spatio-Temporal Transformer for Unsupervised Multi-View 3D Pose Estimation](#) K Dong, K Riou, J Zhu, A Pastor, K Subrin, Y Zhou, X Yun, Y Sun, IEEE Transactions on

Instrumentation and Measurement, 2024

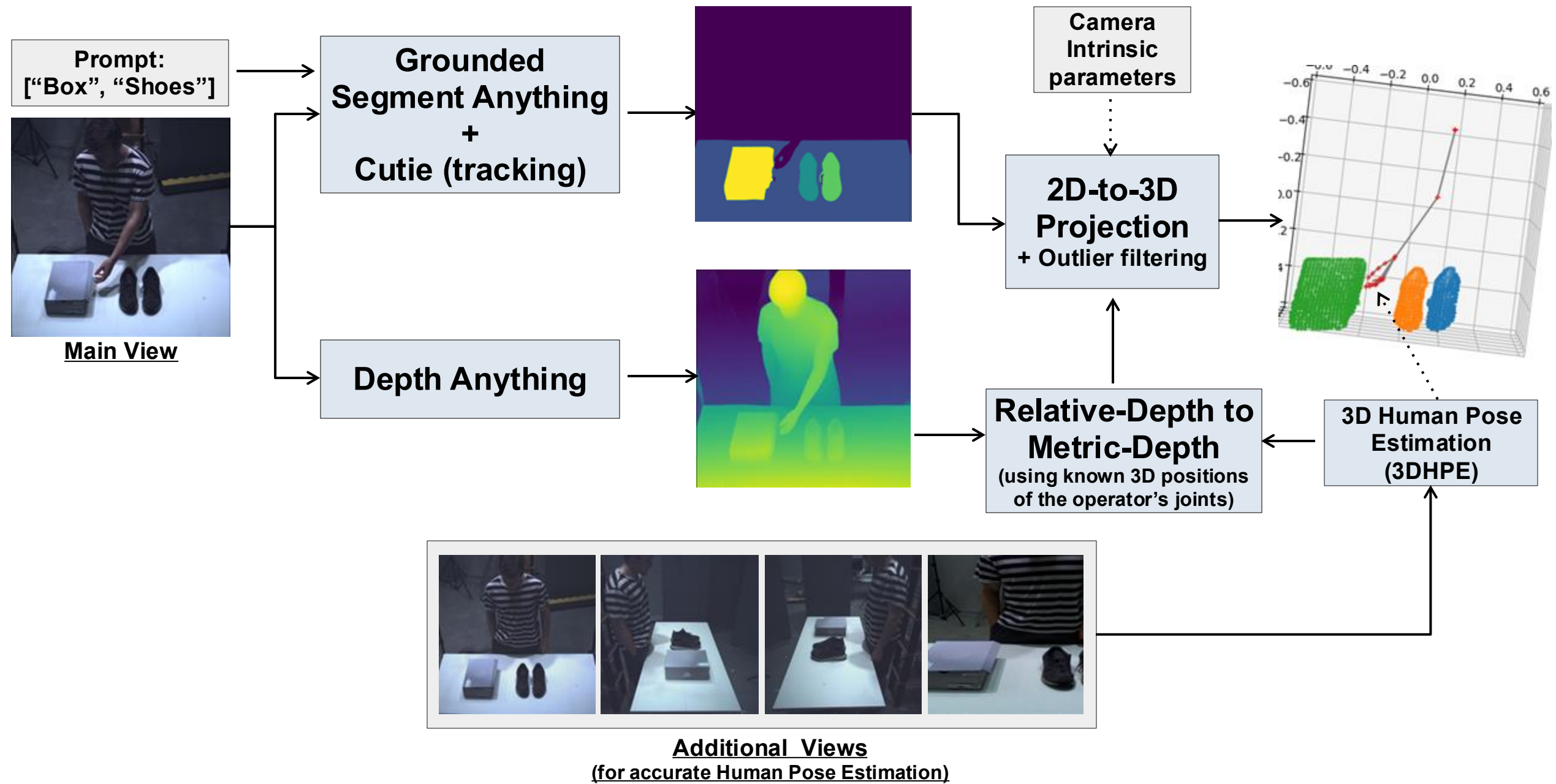


Kinetic particles : from human pose estimation to an immersive and interactive piece of art questionning thought-movement relationships.

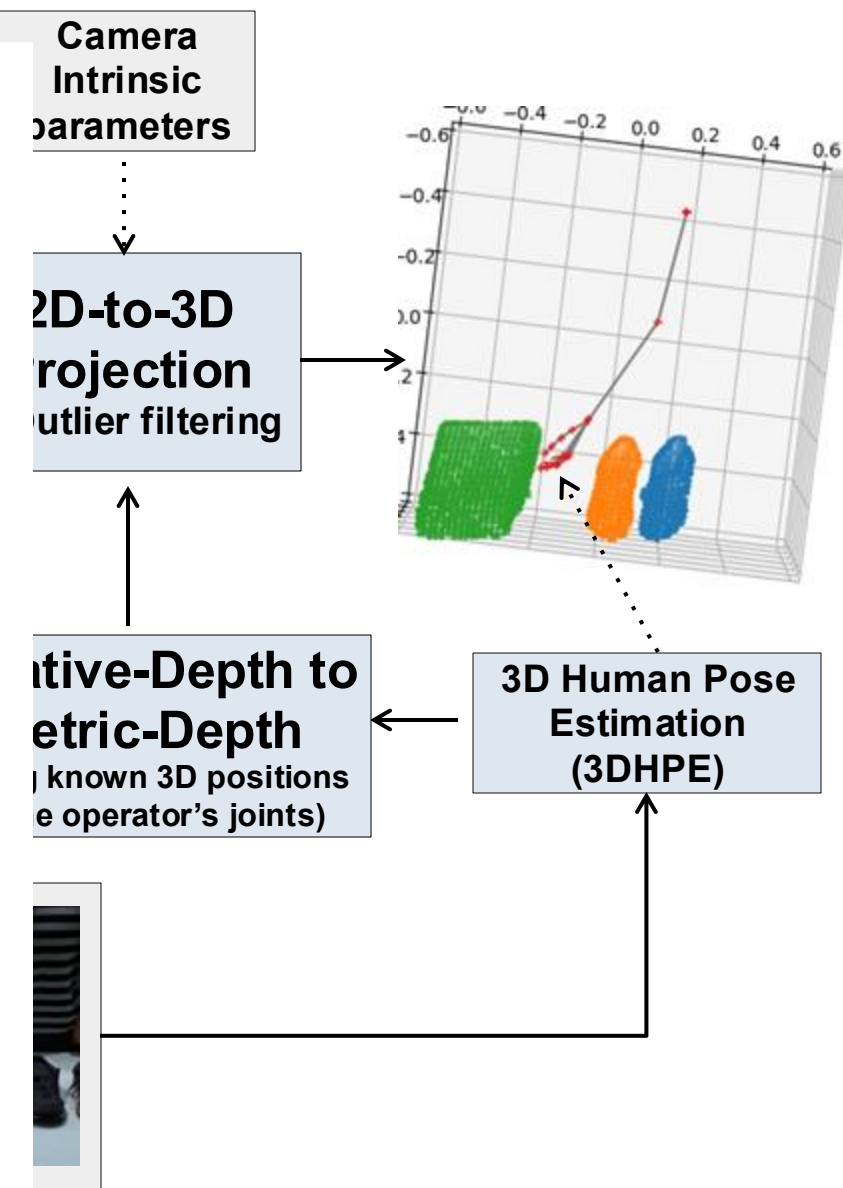
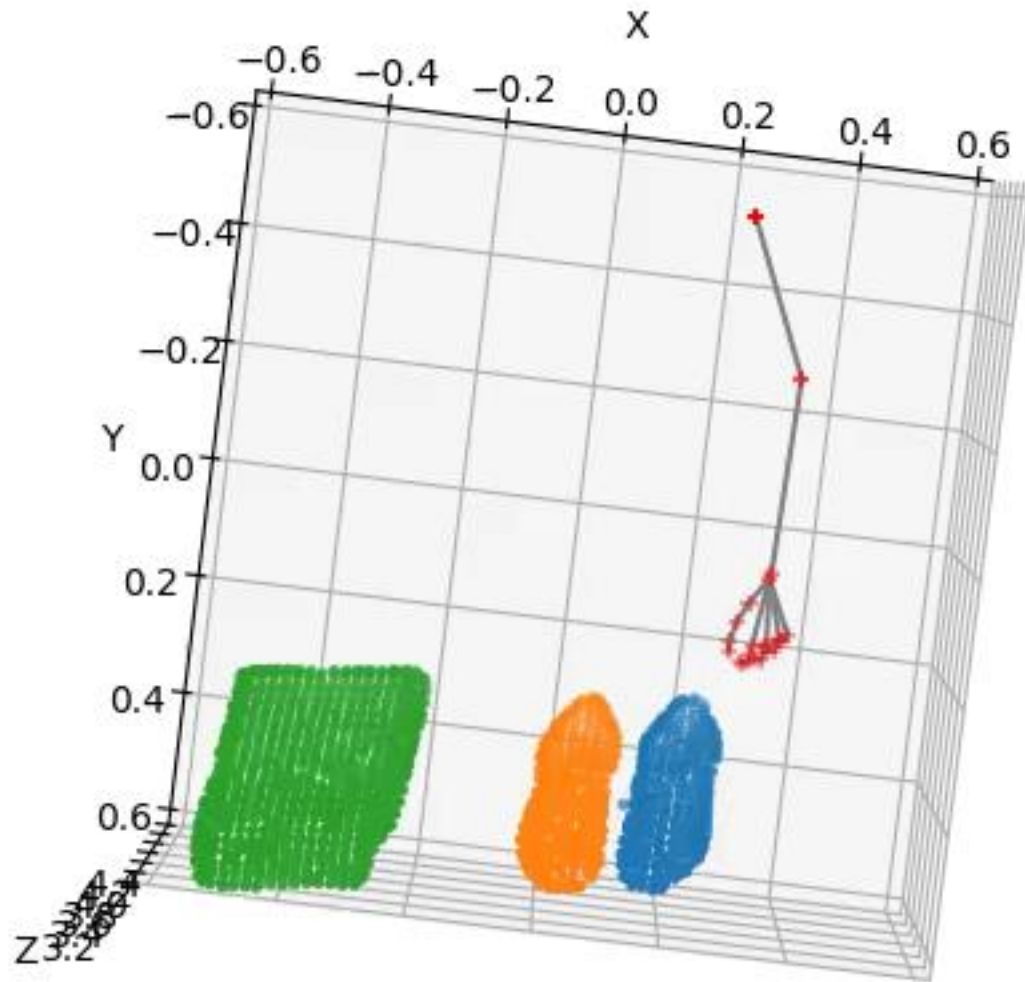
Authors:  [Mickael Lafontaine](#),  [Julie Cloarec-Michaud](#),  [Kevin Riou](#),  [Yujie Huang](#),  [Kaiwen Dong](#), and  [Patrick Le Callet](#) | [Authors Info & Claims](#)

IMX '23: Proceedings of the 2023 ACM International Conference on Interactive Media Experiences • June 2023 • Pages 382 - 385
<https://doi.org/10.1145/3573381.3597228>

(Best Demo award)



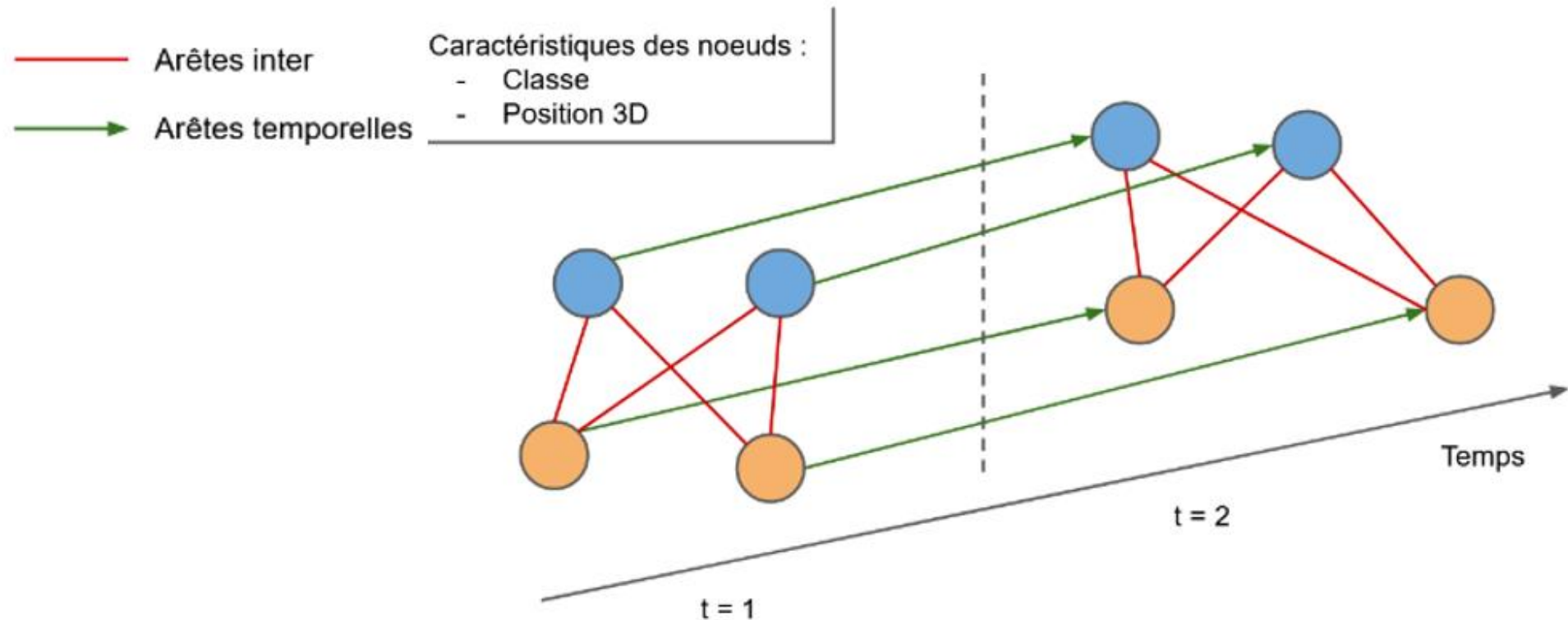
[Vision Foundation Models for an embodiment and environment agnostic scene representation for robotic manipulation](#) K Riou, K Subrin, P Le Callet, IROS 2024





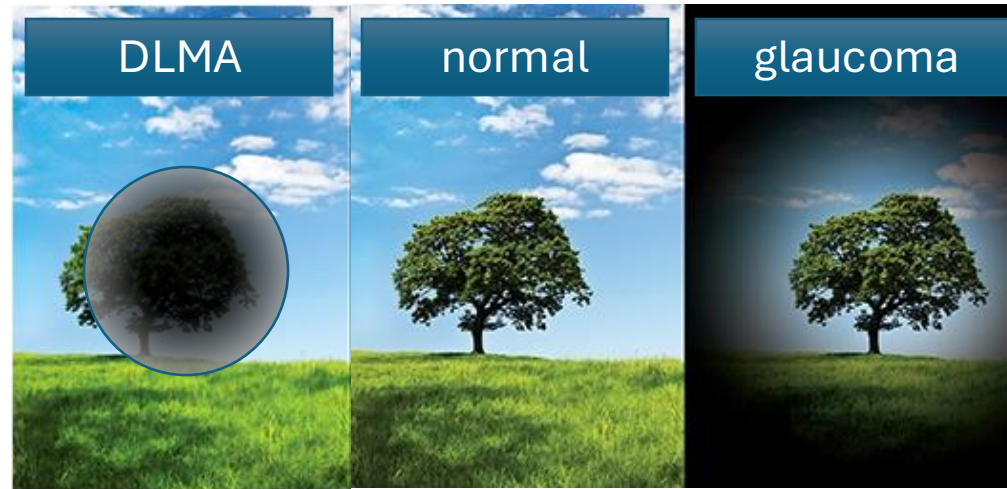
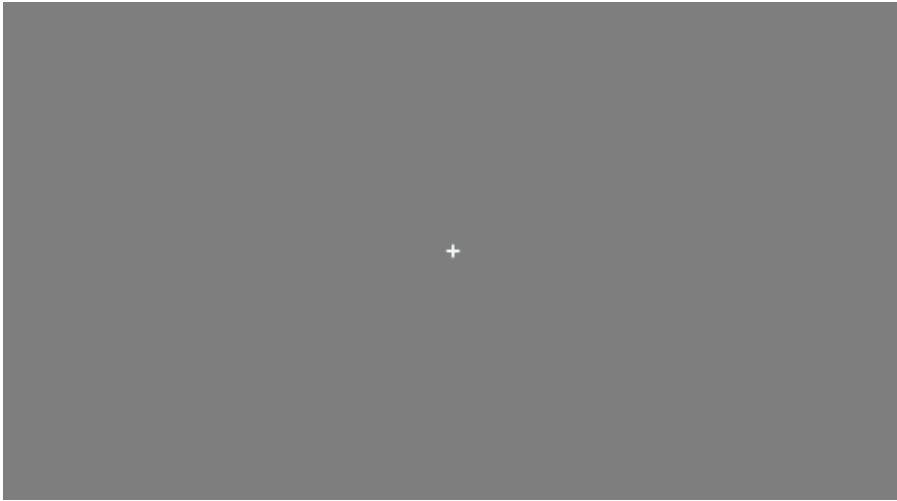


Reconnaissance d'Actions sur Graphes



Détection d'intention?

Gaze contingent display & Patient

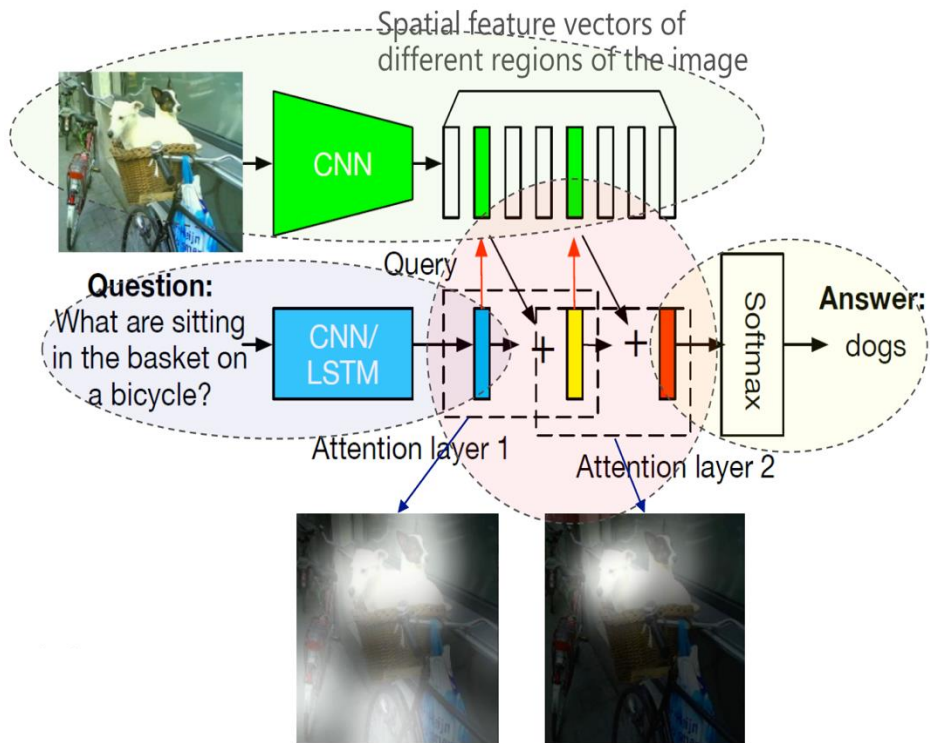


What are the visuo-motor tendencies of omnidirectional scene free-viewing in virtual reality? EJ David, P Lebranchu, MP Da Silva, P Le Callet, Journal of Vision 2022

Predicting artificial visual field losses: a gaze-based inference study E David, P Lebranchu, M Perreira Da Silva, P Le Callet, Journal of Vision 2019

How are ocular behaviours affected by central and peripheral vision loss? A study based on artificial scotomas and gaze-contingent paradigm. E David, P Le Callet, M Perreira Da Silva, P Lebranchu Electronic Imaging 2018

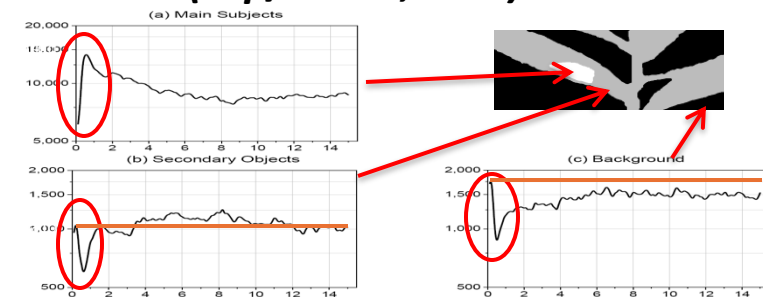
Cognitive computing ? Attentional processes in deep learning algorithms



Foveal (details) vs Periphery (global)

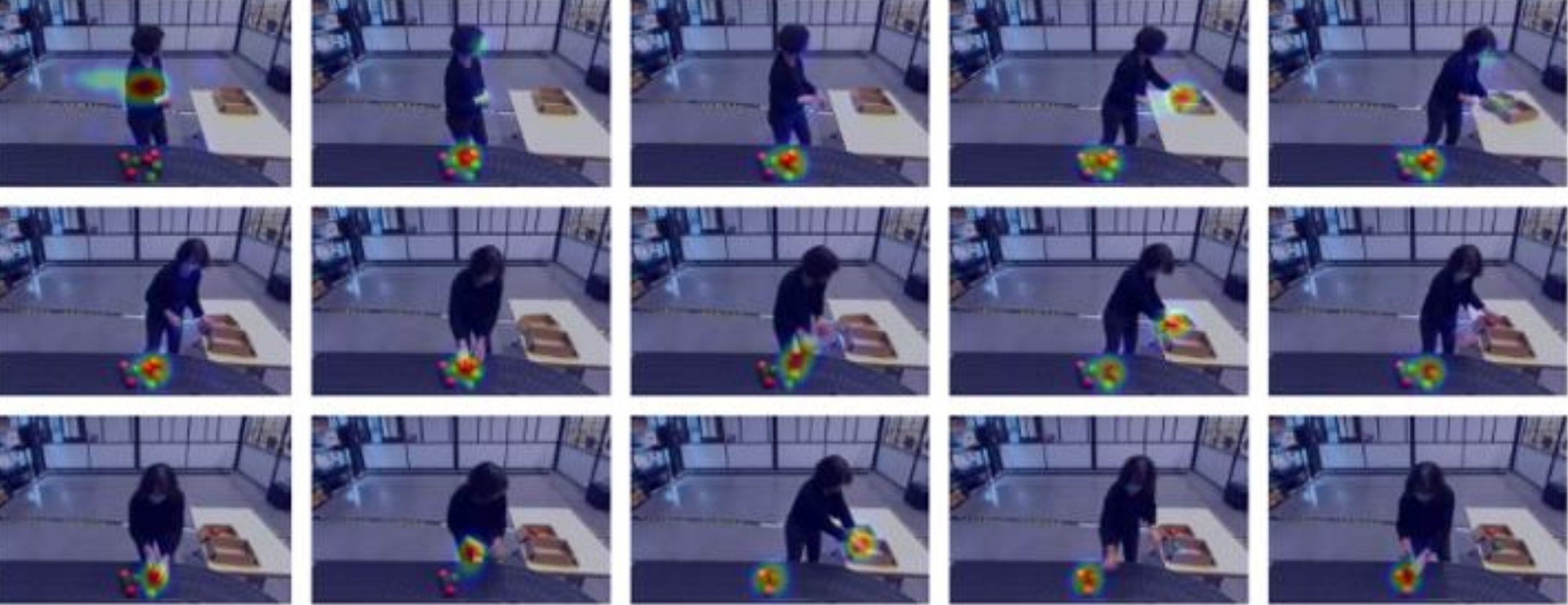


Reflex (bottom up) vs Intentional (top/down, task)



Modeling the visual trace: as annotation or marker of expertise



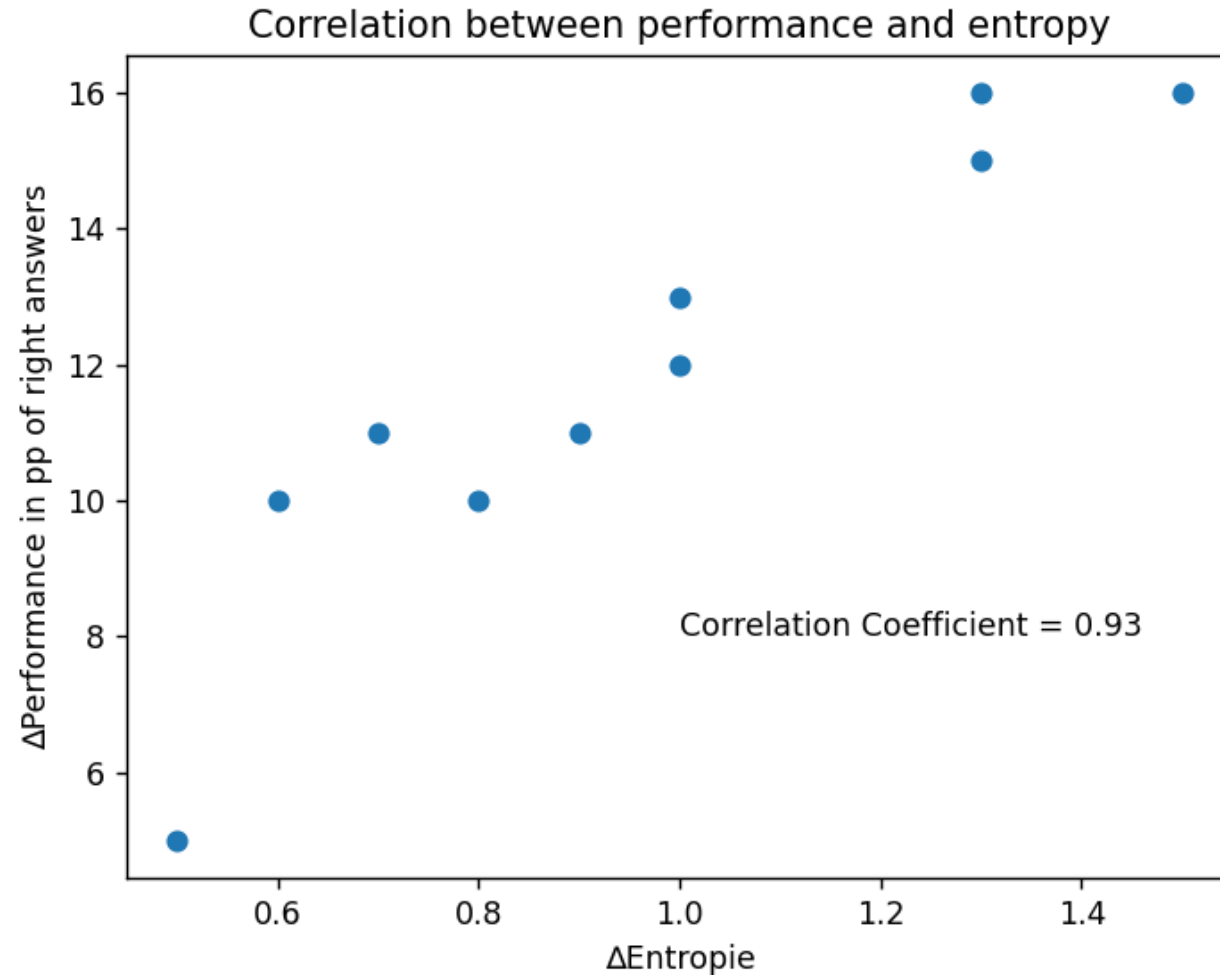


Enjeux/verrous:

- identifier stratégies et nettoyer les données
- explicabilité sur graphe
- LUPI

Numéro du stimulus	Question	Réponse
1	Combien de poivrons rouges sont saisis de la main droite dans cette démonstration ?	2
2	Combien de poivrons rouges sont saisis de la main droite dans cette démonstration ?	4
3	Combien de poivrons verts sont saisis de la main droite dans cette démonstration ?	3
4	Combien de pommes vertes sont saisies de la main droite dans cette démonstration ?	1

Analyse de stratégies visuelles avec HMM (et clustering de HMMs) – Badr tahri





End-to-End Human-Gaze-Target Detection with Transformers

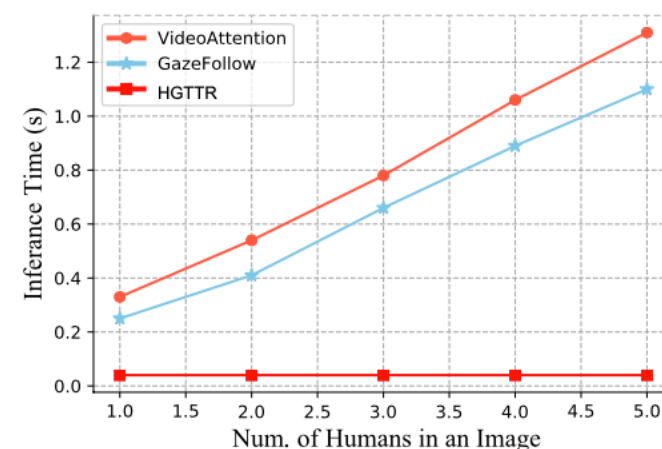
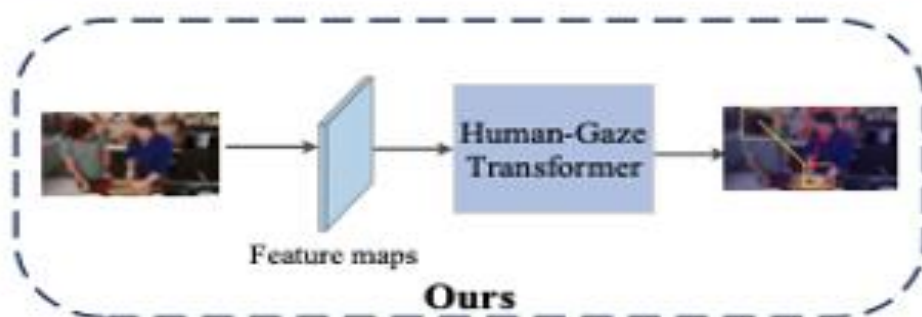
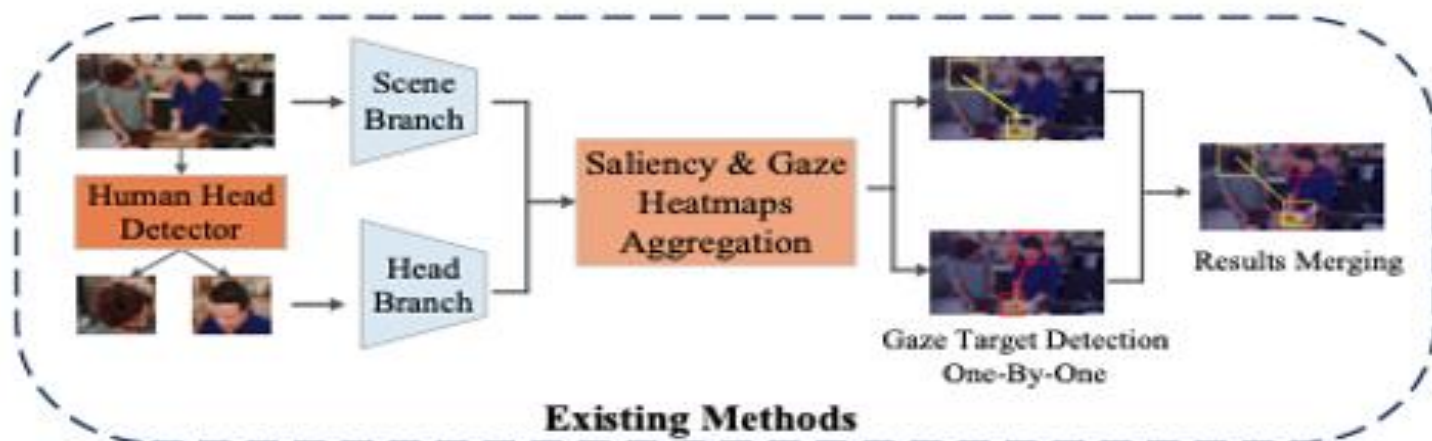
Danyang Tu¹, Xiongkuo Min¹, Huiyu Duan¹, Guodong Guo², Guangtao Zhai^{1*}, Wei Shen^{3*}

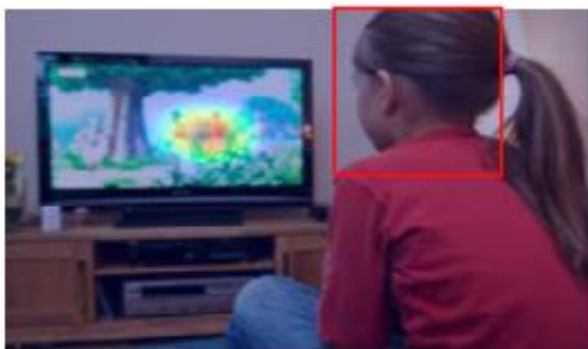
¹Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University

²Institute of Deep Learning, Baidu Research, Beijing, China

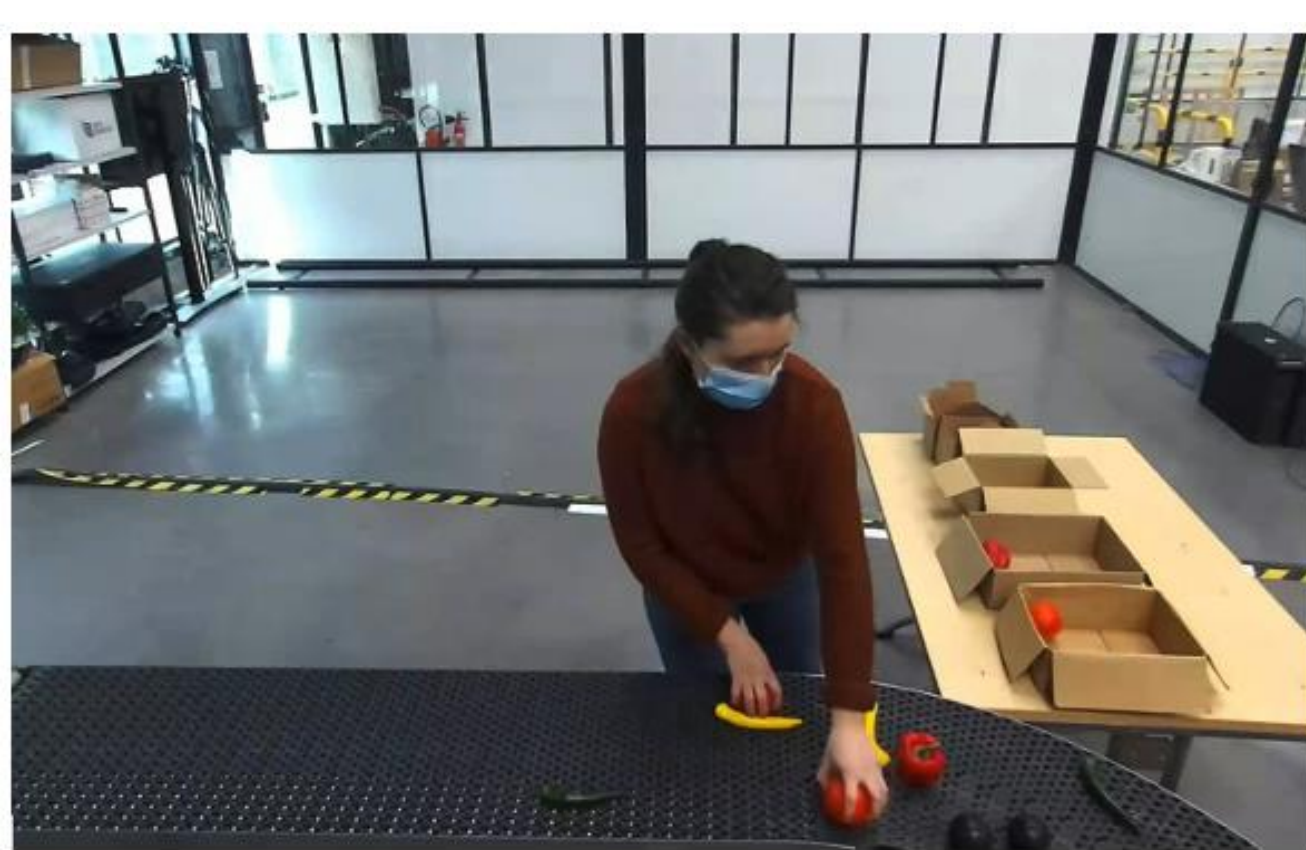
³MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University

{danyangtu, minxiongkuo, huiyuduan, zhaiguangtao, wei.shen}@sjtu.edu.cn, guogudong01@baidu





Et avec des LMMs?



Prompt : Describe this scene. In this scene, what is the woman looking at? to which object does the woman's gaze point towards?

Llava output: *In the image, a woman is standing in a room, wearing a face mask and holding a banana. She is surrounded by several other people, some of whom are also wearing face masks. **The woman's gaze is directed towards a box of oranges**, which is placed on a table nearby. **The other people in the room are also looking at the box of oranges**, indicating that it might be an interesting or unusual object in the scene.*

Chat GPT-4o: In this scene, the woman is looking down at the table where she is sorting vegetables. **Her gaze is directed towards the red bell pepper she is holding in her right hand.** The red bell pepper is the object her gaze points towards.



Prompt: Describe the scene, detect all visible human heads, and for each head, tell me which object does the person's gaze point towards?

Llava output: In the image, there are two men standing in a room, with **one of them looking at a computer screen**. The man looking at the computer screen is wearing a blue shirt, while the other man is wearing a white shirt. **Both men are holding cell phones** in their hands.

There are a total of three visible human heads in the scene. The first head belongs to the man in the blue shirt, who is looking at the computer screen. The second head belongs to the man in the white shirt, who is standing next to the man in the blue shirt. The third head is not visible in the image.

Chat GPT-4o: ### Objects Each Person is Looking At:

- ****Man on the Left**** : ****Gaze Points Towards****: The smartphone in his hands.

****Man on the Right**** : ****Gaze Points Towards****: The computer monitor he is seated in front of.

Overall, the scene captures a moment where the man on the left is focused on his smartphone, and the man on the right is engaged with his computer screen



Prompt: Describe the scene, detect all visible human heads, **describe all objects in the scene**, and for each head, tell me which specific object in this scene does the person's gaze point towards?

Chat GPT-4o: ### Objects Each Person is Looking At:

- ****Man on the Left**** : ****Gaze Points Towards****: **The laptop on the desk in front of him**
- **Man on the Right**** : ****Gaze Points Towards****: **The whiteboard on the wall to his left**

Objects in the Scene:

- ****Laptop****: Positioned in front of the man on the left.
- ****Yellow Coffee Mug****: With the text "Pão de milho" on it, placed on the desk near the man on the left.
- ****Pen Case****: **There are two cases** on the desk; one is black, and the other is red and black.
- ****Remote Control****: Positioned on the desk in front of the man on the right.
- ****Cables****: There are a few cables on the desk, likely connected to the laptop or other devices.



cerveaux fatigués



institut
universitaire
de France



Nantes
Université



L'IA est une
problématique RH



Le LaborIA

Un laboratoire de recherche-action dédié à l'impact de l'intelligence artificielle dans le milieu professionnel.



1. Ce que l'intelligence artificielle fait de l'homme au travail. Visite sociologique d'une entreprise

Yann Ferguson

DANS **LES MUTATIONS DU TRAVAIL (2019)**, PAGES 23 À 42



Chiffres clés

64%

des individus utilisent l'IA dans le cadre de leurs activités professionnelles

Selon une étude de la MIT Sloan Management Review et du cabinet BCG sur la valeur ajoutée individuelle et organisationnelle de l'IA (2022)

60%

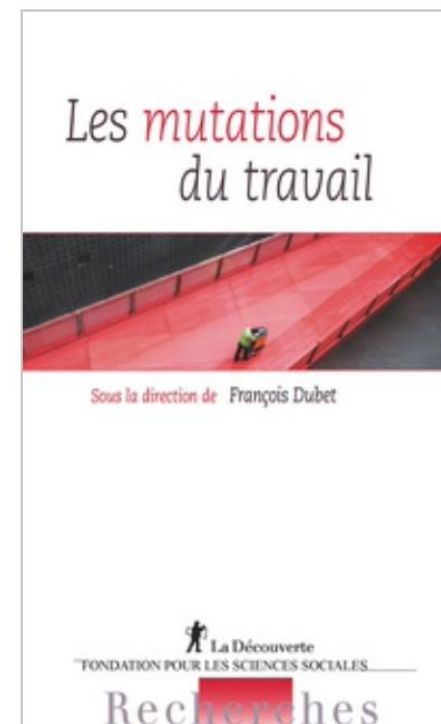
des individus considèrent l'intelligence artificielle comme un partenaire de travail

Selon une étude de la MIT Sloan Management Review et du cabinet BCG sur la valeur ajoutée individuelle et organisationnelle de l'IA (2022)

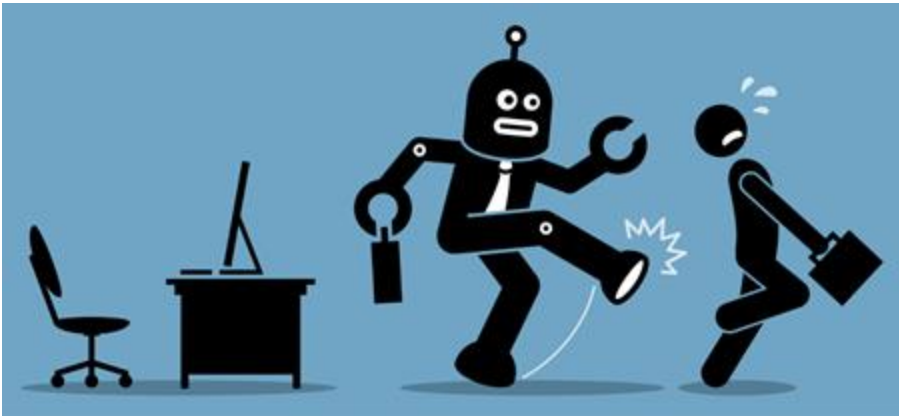
62%

des français pensent que l'IA au travail est une menace pour les salariés et entreprises

Selon l'enquête "Intelligence artificielle au travail" de l'Observatoire de la Tech (2021)



Pourquoi IA & Acceptabilité ?



*Soumission à l'IA,
Remplacement par l'IA*

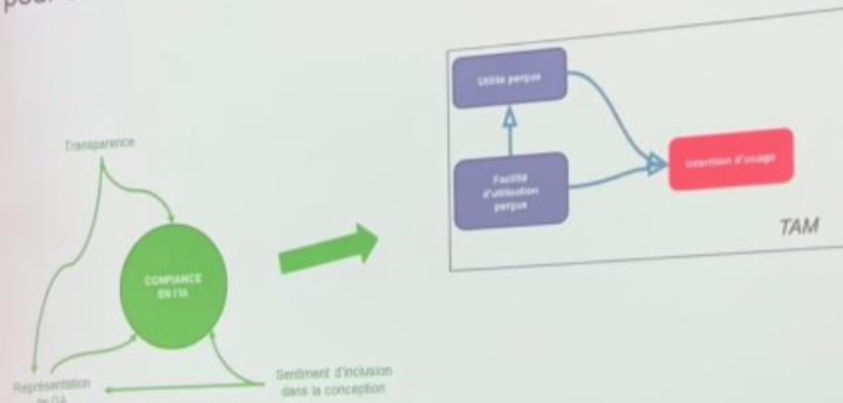


*Augmentation par l'IA,
Assistance par l'IA*

Liens entre confiance et acceptabilité dans un dispositif IA Agossah, L Lévêque, MP da Silva, P Le Callet, F Krupa, G Deconde, IHM 2022

Accroître la prise en compte des utilisateurs
pour concevoir l'IA de demain ?

SI@MA



ÉCOLE DOCTORALE N° 641
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *Informatique*

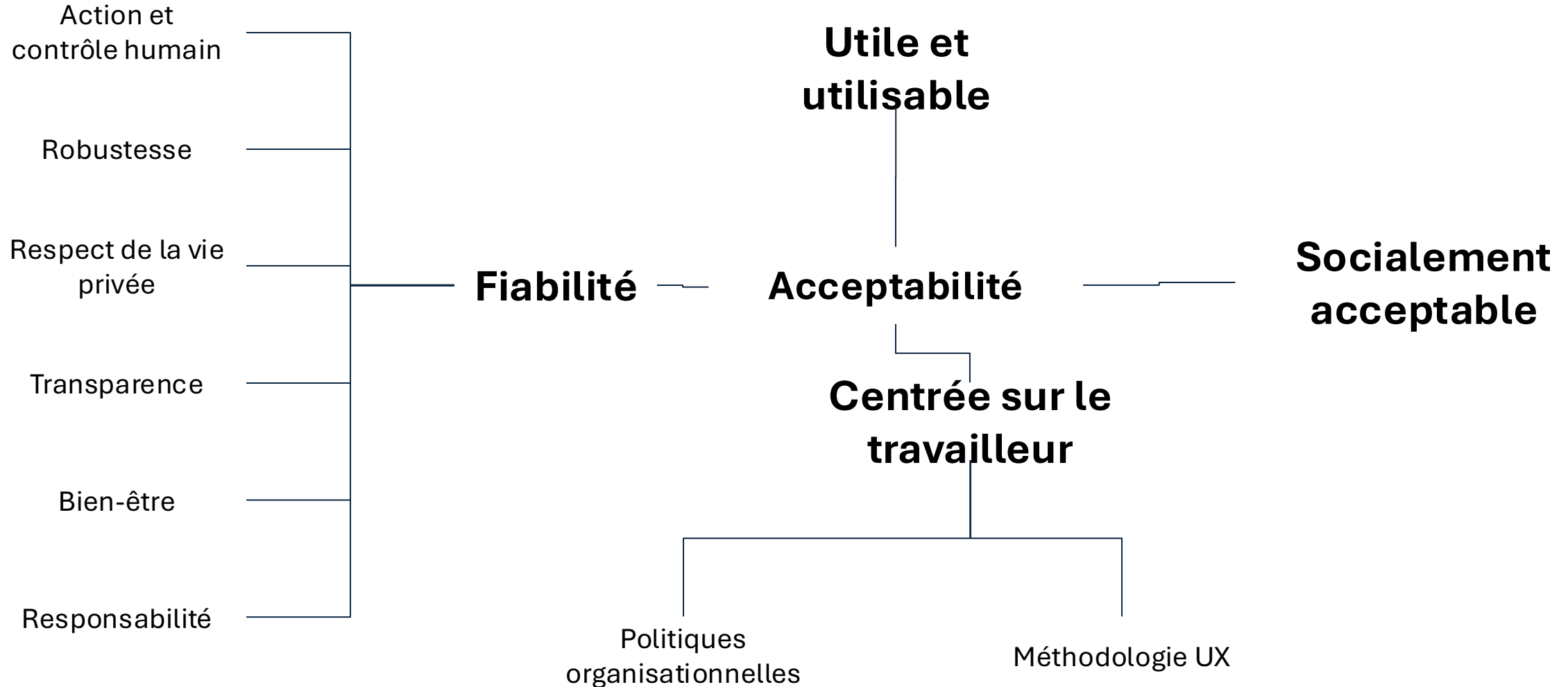
Par

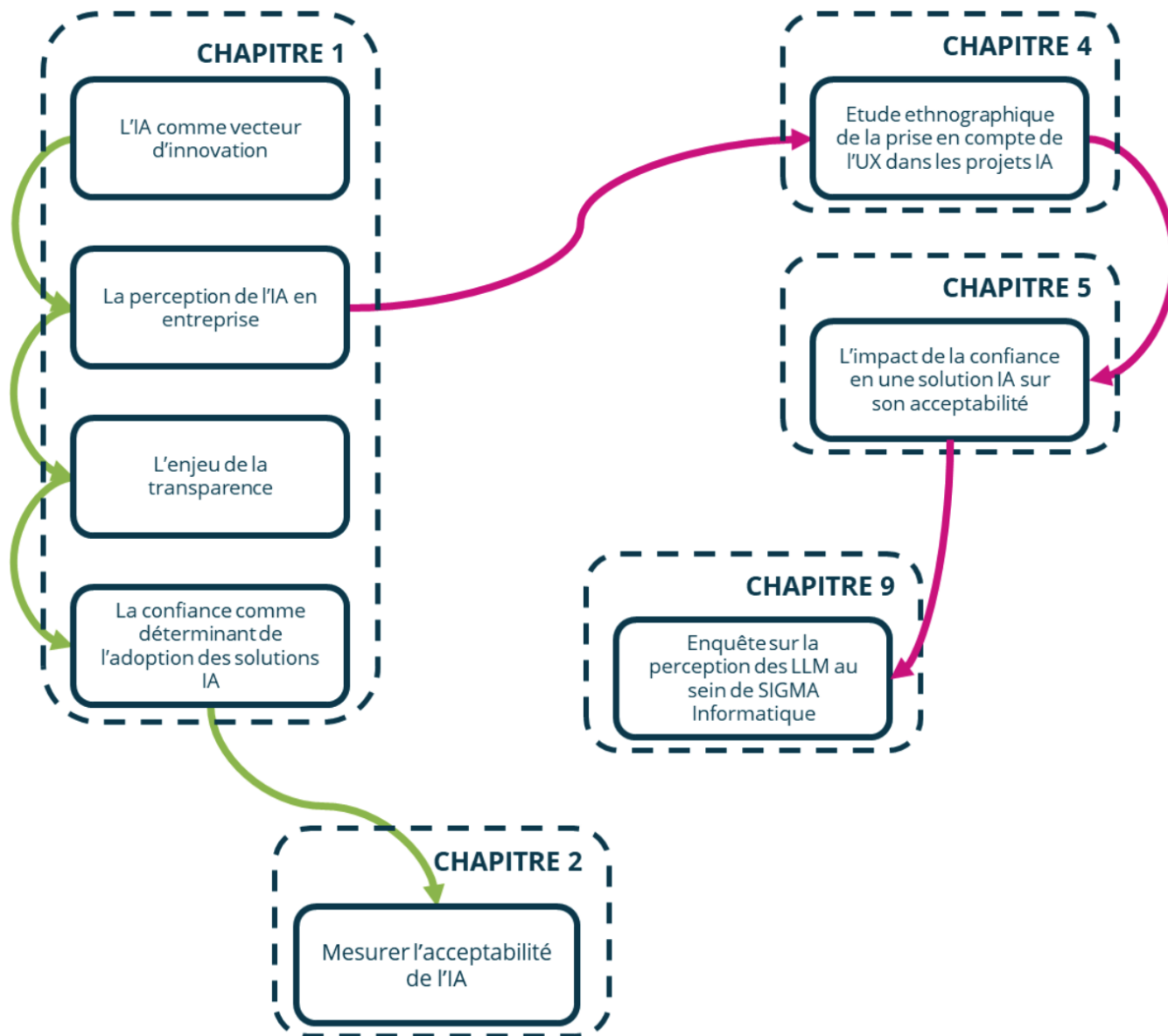
Alexandre AGOSSAH

**Acceptabilité de l'Intelligence Artificielle en contexte
professionnel : Facteurs d'influence et méthodologies
d'évaluation**



Qu'est-ce que c'est une IA acceptable ?





Quelles méthodes pour mesurer de la confiance en des solutions IA ?

Quel impact de la transparence sur la collaboration Humain-IA ?

Quelles typologies de collaboration Humain-IA en contexte professionnel ?

« Pour évaluer cette confiance, nous utilisons notamment des méthodes utilisées pour faire de l'évaluation subjective de la qualité d'expérience (QoE) »

EFFET DE LA CONFIANCE DÉCLARÉE D'UN OUTIL PRÉDICTIF SUR LA CONFIANCE ACCORDÉE PAR LES UTILISATEURS



PRÉDICTION : 31 ANS

CONFIANCE : FAIBLE

Pensez-vous que cette personne est dans la tranche d'âge 25 – 32 ans ?

☒ Oui

☐ Non

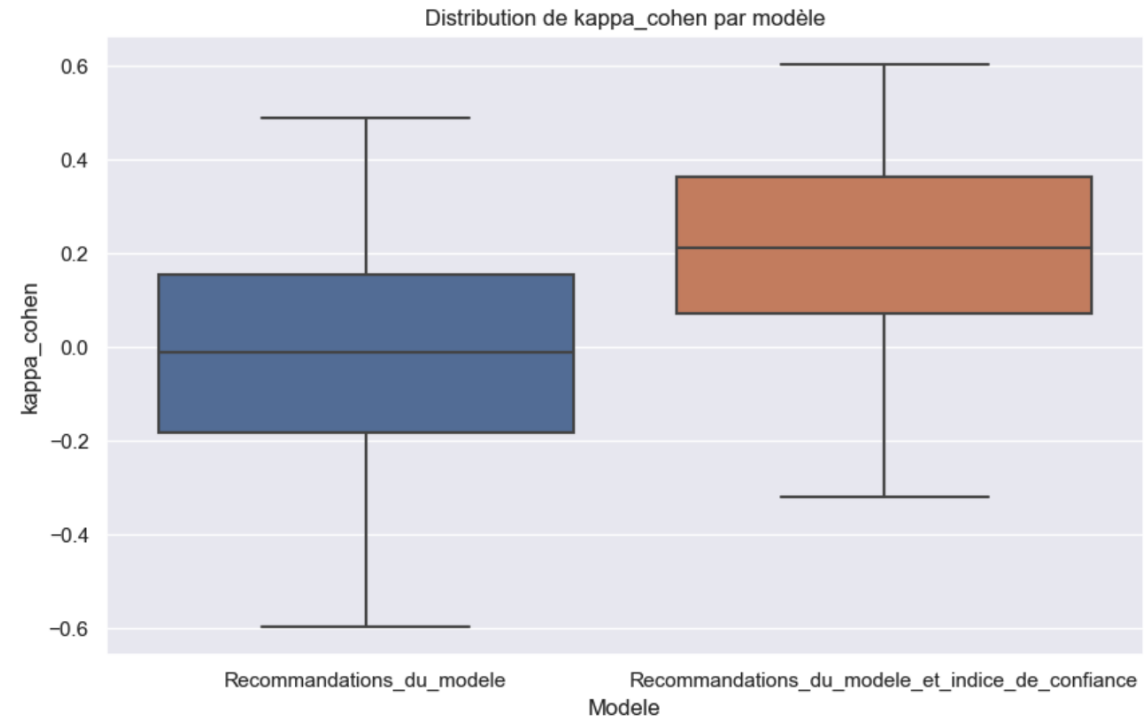


FIGURE 5.10 – Distribution des scores Kappa de Cohen pour les accords H-IA en fonction des conditions expérimentales

L'émergence des outils génératifs : des facteurs d'acceptabilité différents ?

COMMUNICATIONS CACM.ACM.ORG OF THE ACM 03/2024 VOL.67 NO.03

Measuring GitHub Copilot's Impact on Productivity

Experienced developers who are already highly skilled are less likely to write better code with Copilot, but Copilot can assist their productivity in other ways.

until mornings 7:00 am PST, where the average acceptance rate is also rather high at 23%.

• Typical working hours were from 7:00 am PST to 4:00 pm PST, where the average acceptance rate is much lower at 21.2%.

Conclusions

When we set out to connect the productivity benefit of GitHub Copilot to usage measurements from developer activity, we collected measurements about acceptance of completions in line with prior work, but also developed persistence metrics, which arguably capture sustained and direct impact on the resulting code. We were surprised to find acceptance rate (number of acceptances normalized by the number of shown completions) to be better correlated with reported productivity than our measures of persistence.

In hindsight, this makes sense. Coding is not typing, and GitHub Copilot's central value lies not in being the way users enter most of their code. Instead, it lies in helping users to make the best progress toward their goals. A suggestion that serves as a useful template to tinker with may be as good or better than a perfectly correct (but obvious) line of code that only saves the user a few keystrokes.

This suggests that a narrow focus on the correctness of suggestions would not tell the whole story for these kinds of tooling. Instead, one could view code suggestions inside an IDE as

making
aster when
an unfav-
age. These
uggest that
developers
eady highly
less likely
opilot, but
ductivity in
hen engag-
automating

only report
they also
ions. How-
ved in the
ceived Pro-
o differing
he connec-
experience

productiv-
t the accep-
s influence
y increases
model or un-
techniques
these con-
(the study
s to either),
ed temporal

ltural impli-
d weekdays,
as restricted
(whether in
ed the same
estigation in
the absence
acceptance

les collaborateurs de SIGMA Informatique et les outils génératifs ?

Modele TAM + dimension modele UTAUT2

(hypothese: l'émergence récente des outils génératifs, l'attrait pour l'innovation et l'environnement social peuvent être déterminants dans l'intention d'usage

PU – Utilité perçue

PEU - Facilité d'utilisation perçue

SI - Influence sociale

HM – Motivation hédonique

UI - Intention d'usage

UTAUT2: V. Venkatesh and al, « Unified theory of acceptance and use of technology : A synthesis and the road ahead », 2016

TAM: F. Davis, « Perceived Usefulness, Perceived Ease Of Use, And User Acceptance of Information Technology », 1989

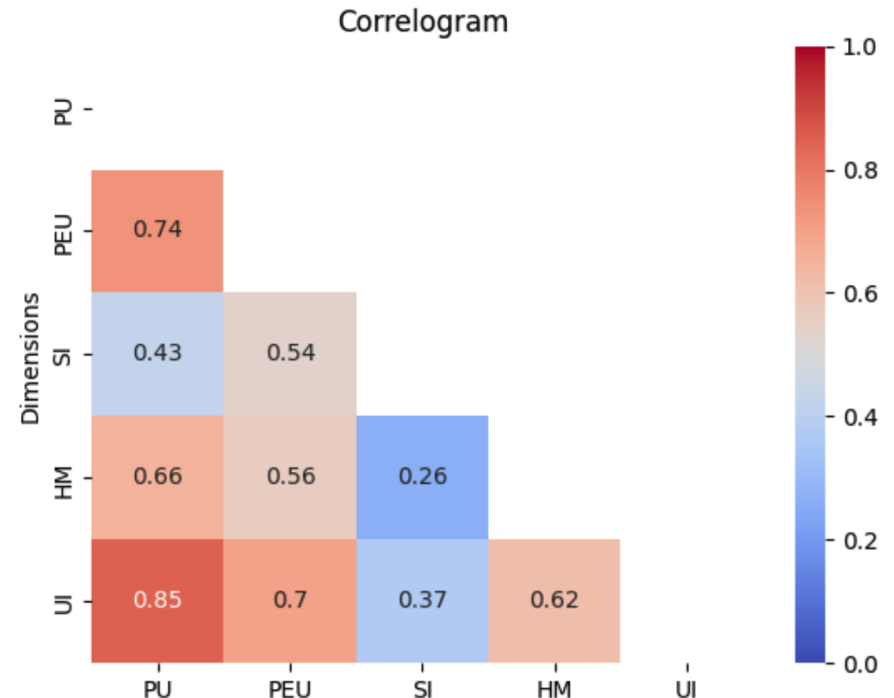
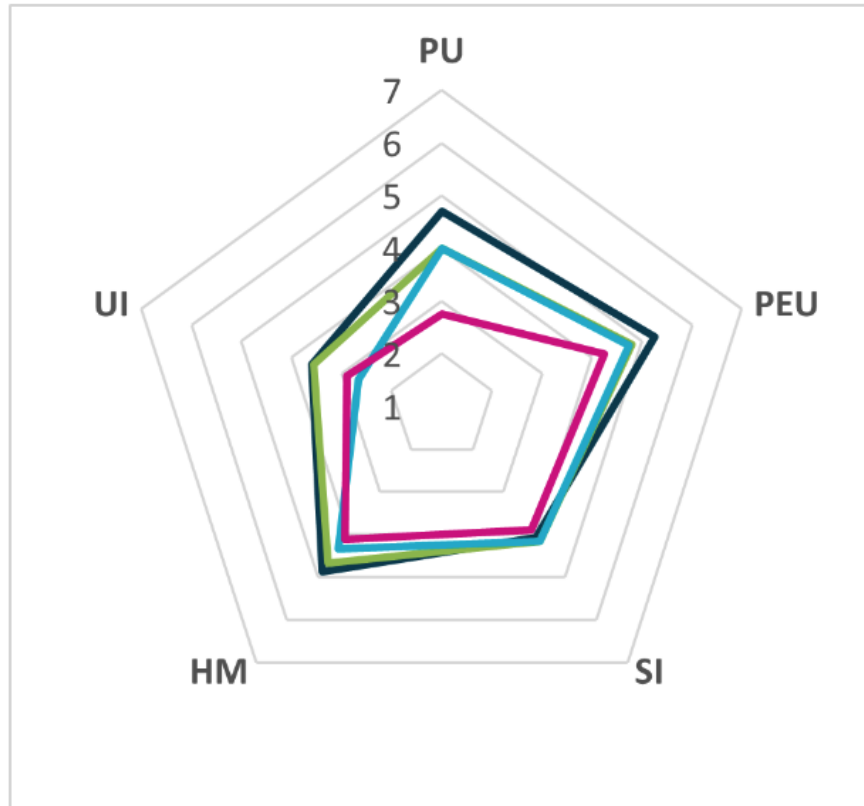
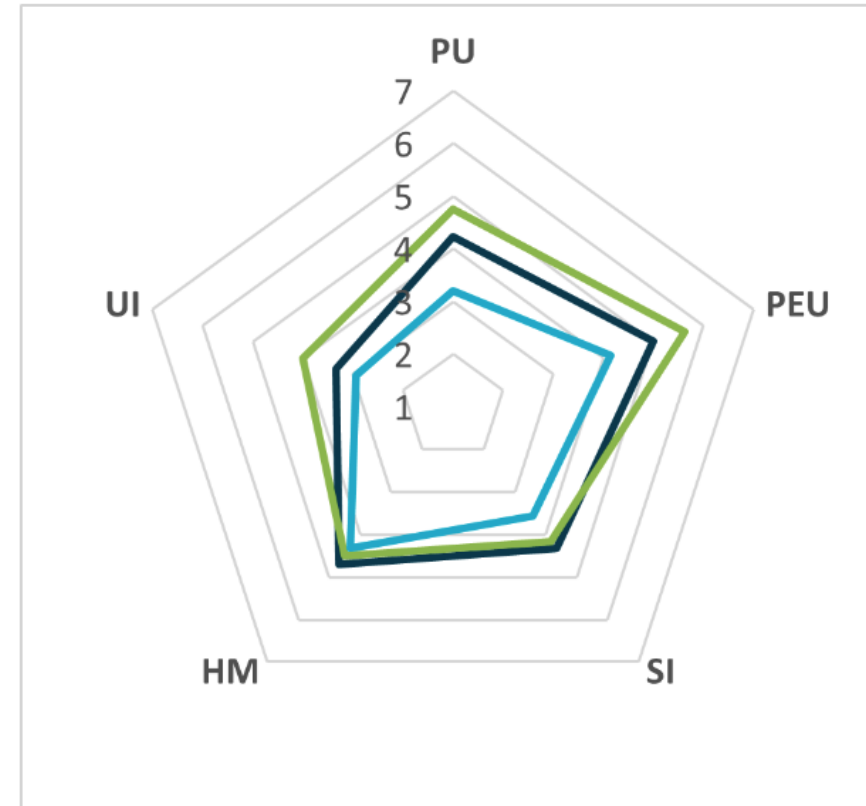


FIGURE 9.3 – Coefficients de corrélation de Pearson entre les dimensions étudiées dans la partie acceptabilité du questionnaire (Légende : plus la valeur r tend vers le rouge plus la corrélation entre les deux dimensions est élevée, à l'inverse plus la valeur r tend vers le bleu plus la corrélation entre les deux dimensions est faible)

les collaborateurs de SIGMA Informatique et les outils génératifs

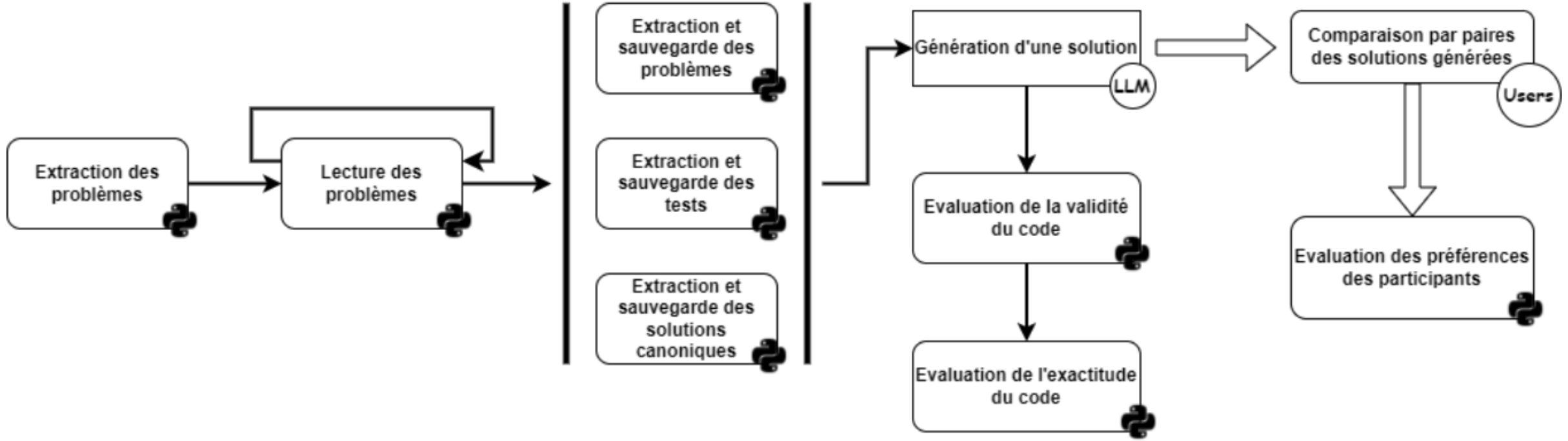


(a) Selon l'expérience en poste (Légende : bleu foncé - moins de 2 ans d'expérience, vert - entre 2 et 5 ans d'expérience, bleu clair - entre 6 et 10 ans d'expérience, violet plus de 10 ans d'expérience)



(b) Selon le profil métier (Légende : bleu foncé - développeurs, vert - Commerce, bleu clair - chefferie de projet)

QUEL LLM POUR METTRE EN PLACE UN ASSISTANT DE DÉVELOPPEMENT PYTHON?



LLMs:

GPT-4, Code LLama et Mistral 7B

17 développeurs Python/ 3 groupes d'expérience :

- 5 développeurs juniors (moins de 2 ans d'expérience) ;
- 10 développeurs intermédiaires (2 à 6 ans d'expérience) ;
- 2 développeurs seniors (plus de 6 ans d'expérience)

QUEL LLM POUR METTRE EN PLACE UN ASSISTANT DE DÉVELOPPEMENT PYTHON

GPT-4:

- populaire
- score élevé aux datasets d'évaluation

Et bien

nos résultats nous indiquent que les développeurs sont susceptibles de préférer Mistral 7B et Code Llama 7B plus frugaux

« réponses moins fournies et plus concises »



Ressources rares



institut
universitaire
de France



Nantes
Université



Pourquoi notre utilisation de l'intelligence artificielle est un gouffre énergétique

Par Léa Prati et Romain Geoffroy

Publié hier à 05h00, modifié hier à 22h01

🕒 Lecture 6 min.

Lire dans l'application



DÉCRYPTAGE | De la conception d'un modèle à son utilisation, l'IA générative s'avère extrêmement énergivore. « Le Monde » détaille les raisons de cette voracité, alors que les usages sont en pleine expansion.

FRANKENSTREAM

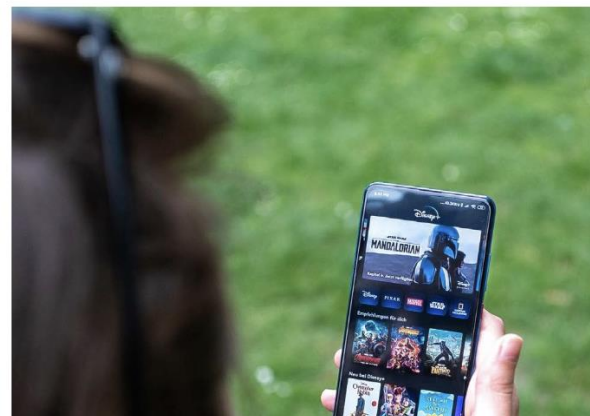
CE MONSTRE QUI NOUS DÉVORE

Agir pour une société positive

Projet Diffusion durable et responsable des contenus vidéo

Objectifs du projet :

- **Adapter** la qualité d'expérience perçue d'une vidéo en fonction d'un contexte d'usage donné (taille du forfait mobile, énergie de la batterie disponible, qualité du signal de réception)
- **Identifier** les écarts de compression des contenus générés par les particuliers (contenus UGC) et définir les efforts d'optimisation à opérer selon la popularité du contenu
- **Adapter** la compression aux encodeurs les plus récents
- **Mesurer** la qualité de vidéo HDR et l'interopérabilité entre mobiles



Une démarche d'innovation ouverte pour une diffusion durable des contenus en streaming

L'aventure de ce projet naît au Laboratoire des sciences du numérique de Nantes (LS2N).

Une équipe de chercheurs reconnue pour son expertise dans le domaine de la perception visuelle des applications numériques travaille sur l'évaluation automatique de la qualité des images et des vidéos et son amélioration.

Une démarche d'open innovation va se mettre en place, fruit d'une collaboration exemplaire entre acteurs industriels et institutions académiques.

Definition of Quality of Experience

Qualinet White Paper on Definitions of Quality of Experience (2012)

P. Le Callet, S. Möller and A. Perkis, eds.

Available at <http://www.qualinet.eu>

⇒ A collective effort of more than 40 authors/contributors (COST Action IC 1003)
European Network on Quality of Experience in Multimedia Systems and Services



Quality of Experience (QoE) is the degree of **delight or annoyance of the user of an application or service.**

It results from the **fulfillment of his or her expectations** with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state.

Quality of Experience: Typology of influencing factors

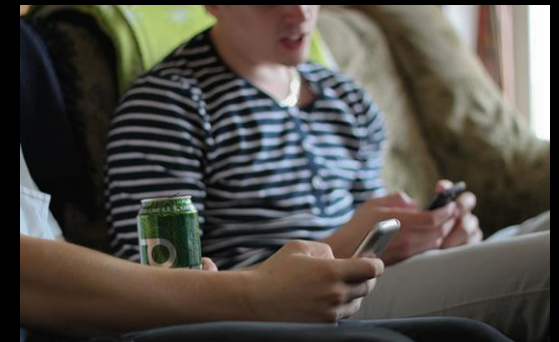
System IF: “properties and characteristics that determine the *technically produced quality* of an application or service. They are related to media capture, coding, transmission, storage, rendering, and reproduction/display, as well as to the communication of information itself from content production to user”



Human IF “any variant or invariant property or characteristic of a human user. The characteristic can describe the *demographic and socio-economic* background, the *physical and mental* constitution, or the user’s *emotional state*”



Context IF “factors that embrace any situational property to describe the *user’s environment* in terms of physical, temporal, social, economic, *task*, and technical characteristics”



Good QoE & Sustainable Development Goals

Human IF: “any variant or invariant property or characteristic of a human user. The characteristic can describe the *demographic and socio-economic* background, the *physical and mental* constitution, or the user’s *emotional state*”



A new momentum for
QoE science

Humans, and Society are
organic

=> shift of expectations



Humans and society are organic: shift of expectations



RÉPUBLIQUE FRANÇAISE
Légifrance

Liberté
Égalité
Fraternité

Le service public de la diffusion du droit

Recommandation n° 2023-02 du 26 juillet 2023 de l'Autorité de régulation de la communication audiovisuelle et numérique relative à l'information des consommateurs, par les services de télévision, les services de médias audiovisuels à la demande et les services de plateformes de partage de vidéos sur la consommation d'énergie et les équivalents d'émissions de gaz à effet de serre de la consommation de données liée à l'utilisation de ces services

NOR : RCAC2324195X
JORF n°0212 du 13 septembre 2023
Texte n° 68



Extrait du Journal officiel
électronique authentifié
PDF - 216 Ko

... HOME / COMMUNICATION ROOM / BLOG / TELECOMS' THIRD WAVE TO DRIVE EU'S GREEN DEAL: AN EFFICIENT USE OF NETWORKS

Telecoms' third wave to drive EU's Green Deal: an efficient use of networks

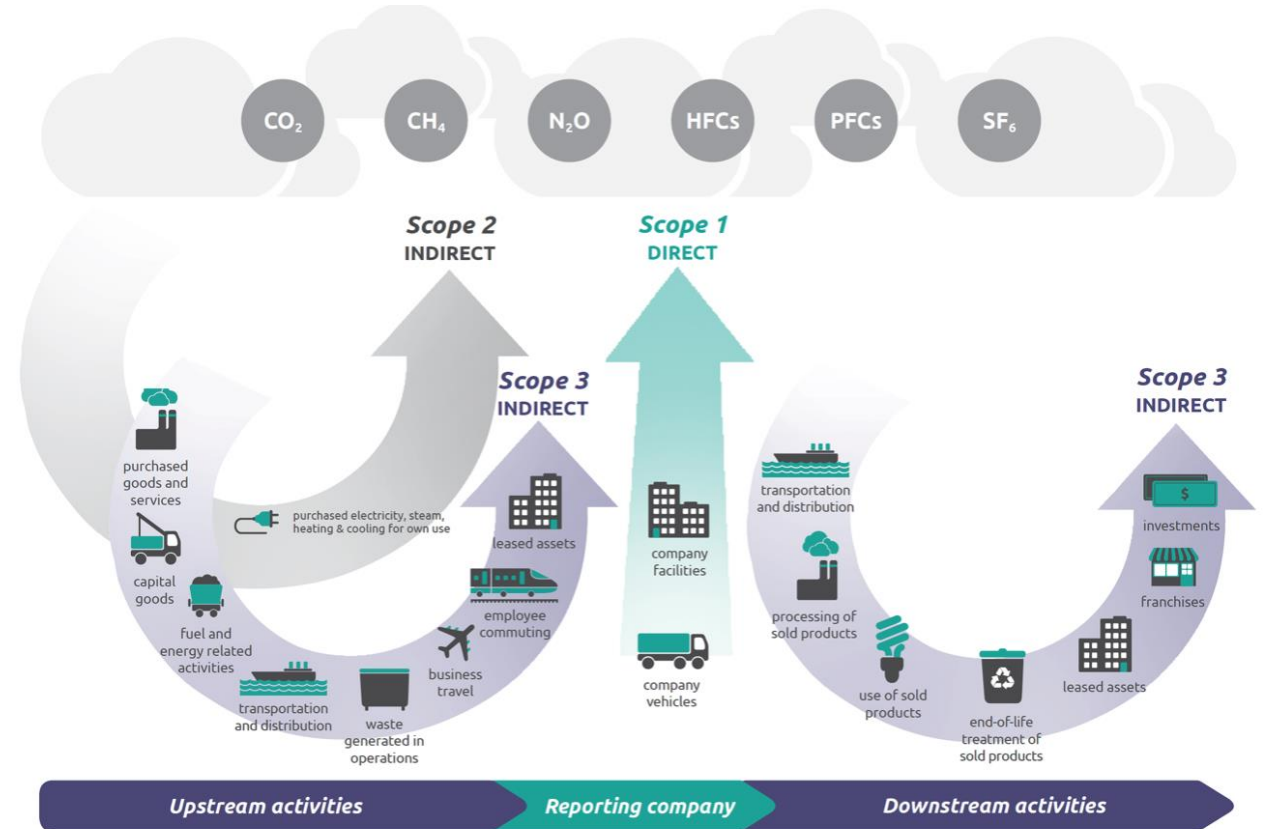
Early October, ARCEP, France's Regulatory Authority, architect and guardian of the country's internet, fixed and mobile networks, launched a public consultation on a reference policy framework for the eco-design of digital services. ARCEP's proposal seeks to provide a reference grid for digital industry stakeholders wanting to commit to sustainable design.

First wave – More efficient network technologies

Second wave – Telecom networks and EU-Taxonomy

Third wave – an efficient use of networks

GHG (Green House Gas) Protocol



Scope 3 emissions are the result of activities from assets not owned or controlled by the reporting organization, but that the organization indirectly affects in its value chain. Scope 3 emissions include all sources not within an organization's scope 1 and 2 boundary. The scope 3 emissions for one organization are the scope 1 and 2 emissions of another organization. Scope 3 emissions, also referred to as value chain emissions, often represent the majority of an organization's total greenhouse gas (GHG) emissions.

AccAnn scale: A QoE scale for Media Quality in Context

- Perceived quality does not necessarily need to directly express the acceptability-annoyance of the video
- Acceptability-Annoyance is a measure of user satisfaction for video streaming services.
- Acceptability-Annoyance takes user expectations and user profile into account ...and possibly with different context

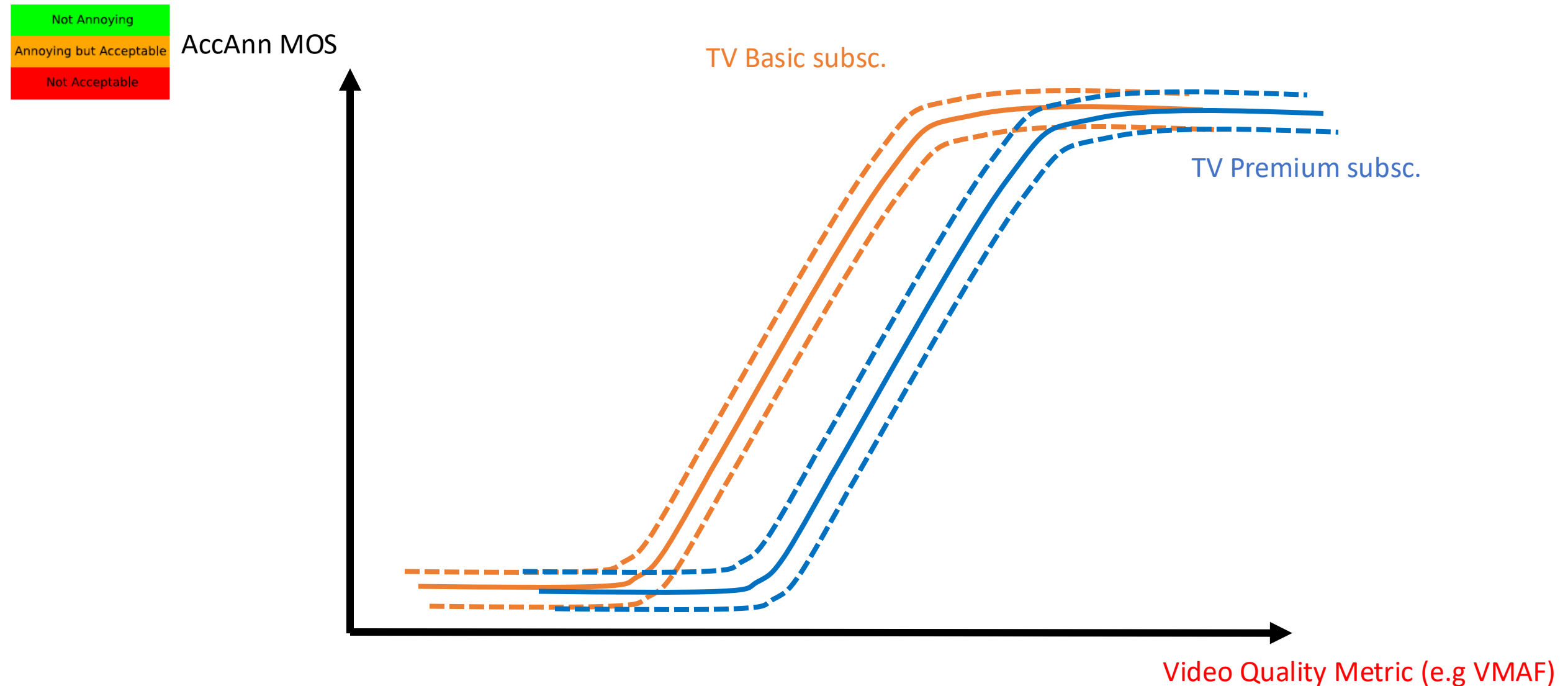
Quality Scale
Absolute Quality
Rating Scale
(ACR)

Excellent
Good
Fair
Poor
Bad

Acceptance-
Annoyance scale
(AccAnn)

Not Annoying
Annoying but Acceptable
Not Acceptable

Mapping Video Quality Metric and AccAnn



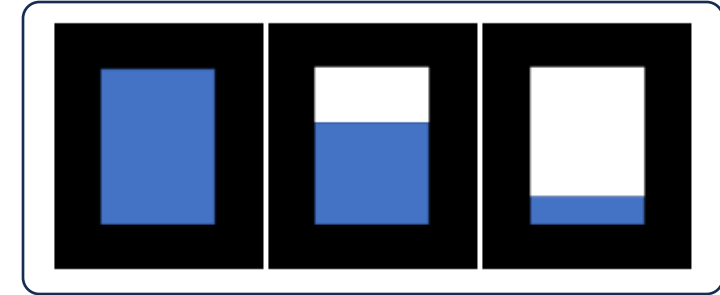
Acceptability & Annoyance and sustainable context



Remaining Data
Remaining Battery
Signal Strength
etc.,

Remaining
Data Plan

High Medium Low

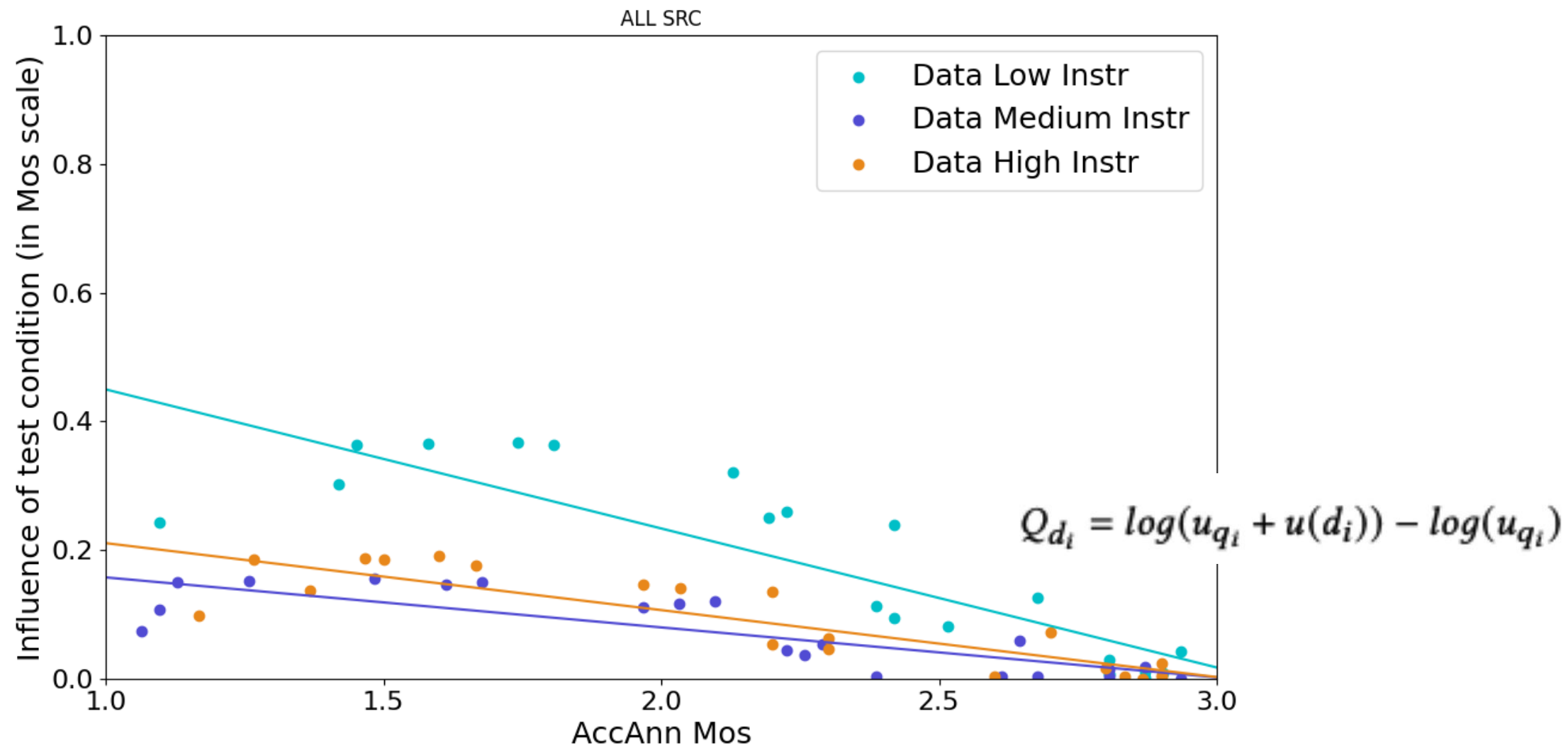


Remaining
Battery



2 - Ali Ak, Anne Flore Perrin, Denise Noyes, Ioannis Katsavounidis, and Patrick Le Callet, "Video consumption in context: Influence of data plan consumption on qoe," in Proceedings of the 2023 ACM International Conference on Interactive Media Experiences, New York, NY, USA, 2023, IMX '23, p. 320–324, Association for Computing Machinery.

Example Outcome of EBA Analysis for Data Context





Une IA pour faire quoi?



institut
universitaire
de France



Nantes
Université





HHS Public Access

Author manuscript

Clin Exp Ophthalmol. Author manuscript; available in PMC 2019 April 01.

Published in final edited form as:

Clin Exp Ophthalmol. 2018 April ; 46(3): 247–259. doi:10.1111/ceo.13022.

Novel mobility test to assess functional vision in patients with inherited retinal dystrophies

Daniel C Chung, DO¹, Sarah McCague, BA², Zi-Fan Yu, ScD³, Satha Thill, MS³, Julie DiStefano-Pappas, BA⁴, Jean Bennett, MD PhD^{2,5}, Dominique Cross, MPH², Kathleen Marshall, COT², Jennifer Wellman, MS¹, and Katherine A High, MD¹



Réalité Virtuelle pour l'évaluation quantitative de la vision fonctionnelle

Chaire Fondamentale de l'Institut Universitaire de France

Pr. Patrick Le Callet



Un frein majeur à l'adoption des biothérapies pour la vision est le manque d'outils pour mesurer objectivement la vision fonctionnelle. Pour démontrer l'impact des traitements prometteurs, nous faisons l'hypothèse que la Réalité Virtuelle (RV), par sa capacité à reproduire des situations écologiques, couplée à une analyse algorithmique avancée de données de navigation, permettrait une évaluation quantitative fine de la vision fonctionnelle pour un suivi longitudinal de l'efficacité clinique



Construire des cohortes de patients et fédérer une communauté proche des usagers



Qualifier les équipements et mener des expérimentations



Diffuser des données ouvertes et stimuler l'innovation ouverte



Penser et développer des environnements écologiques en RV et des protocoles cliniques

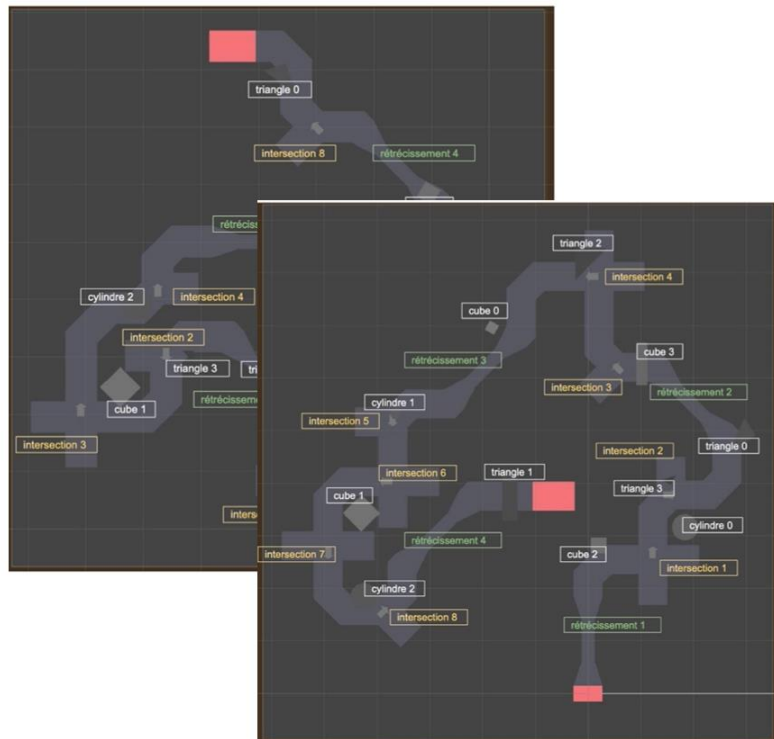
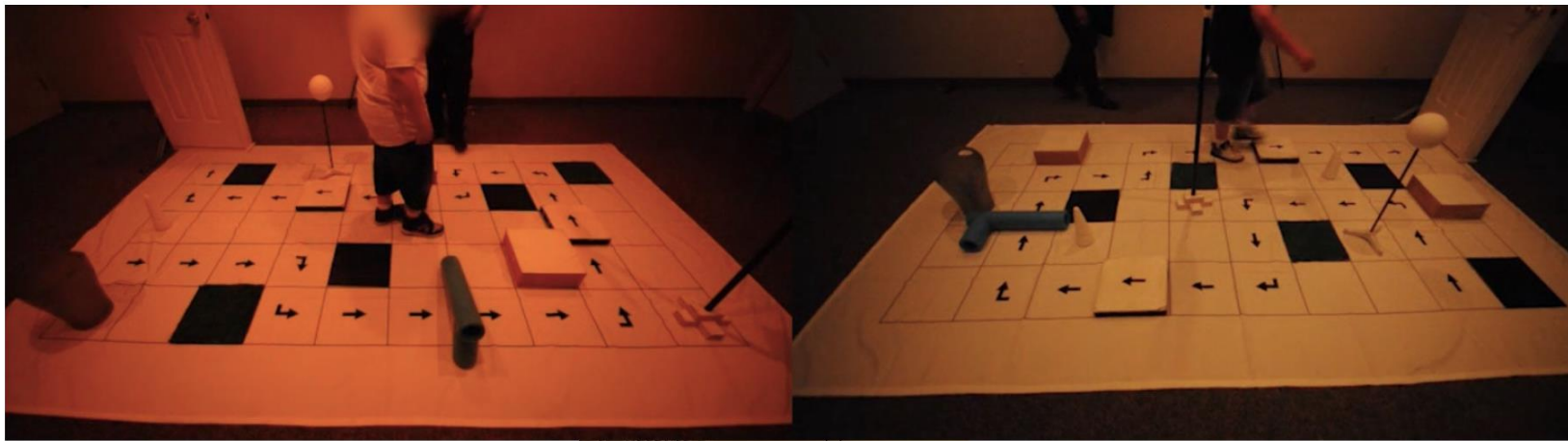


Concevoir des méthodes avancées d'analyse de comportement de navigation et d'interaction



Produire des démonstrateurs de médiation et des mises en situation empathiques

MOVING TO VR



Myopie
forte+glaucome

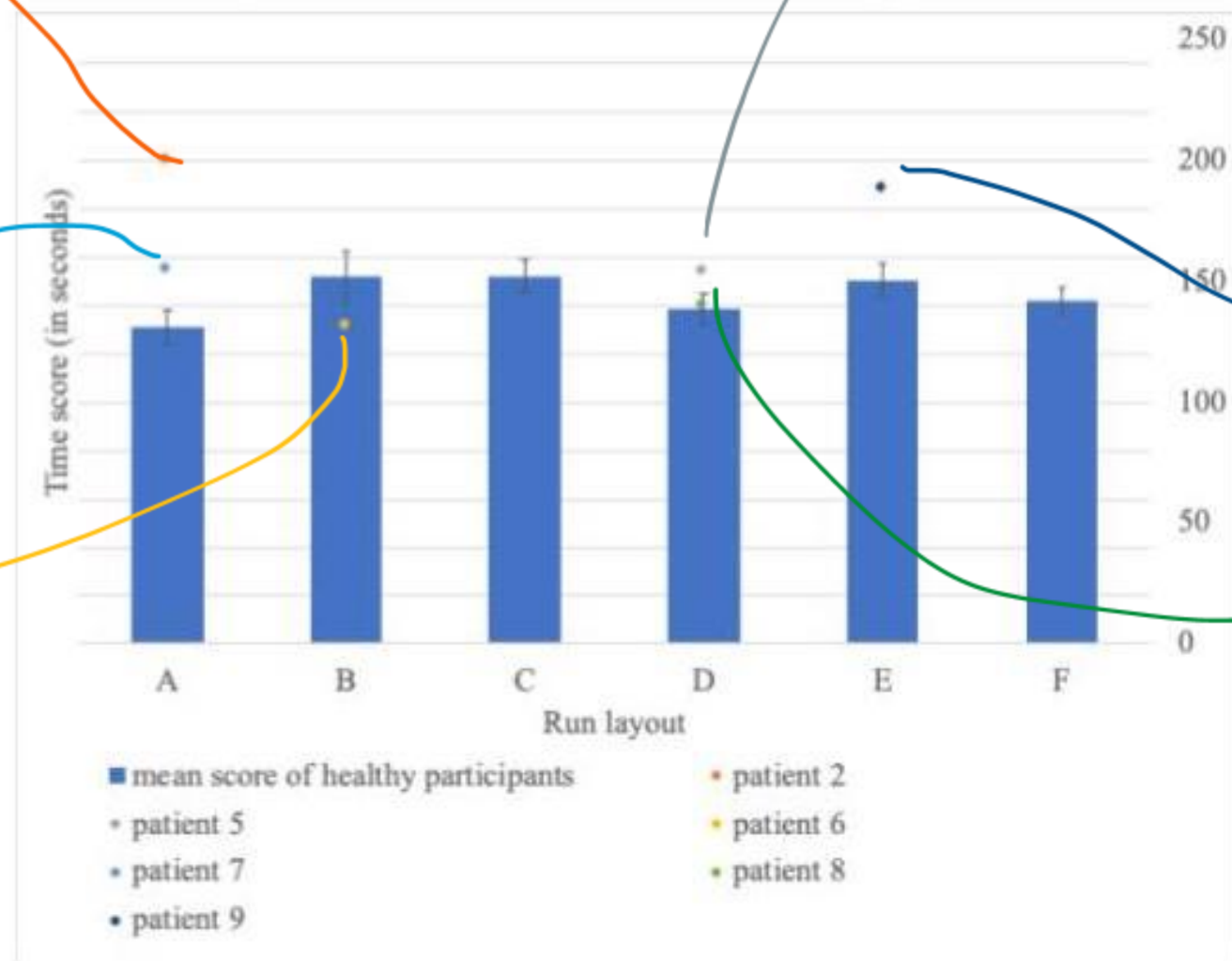
Hémianopsie
bitemporale

Monophtalme

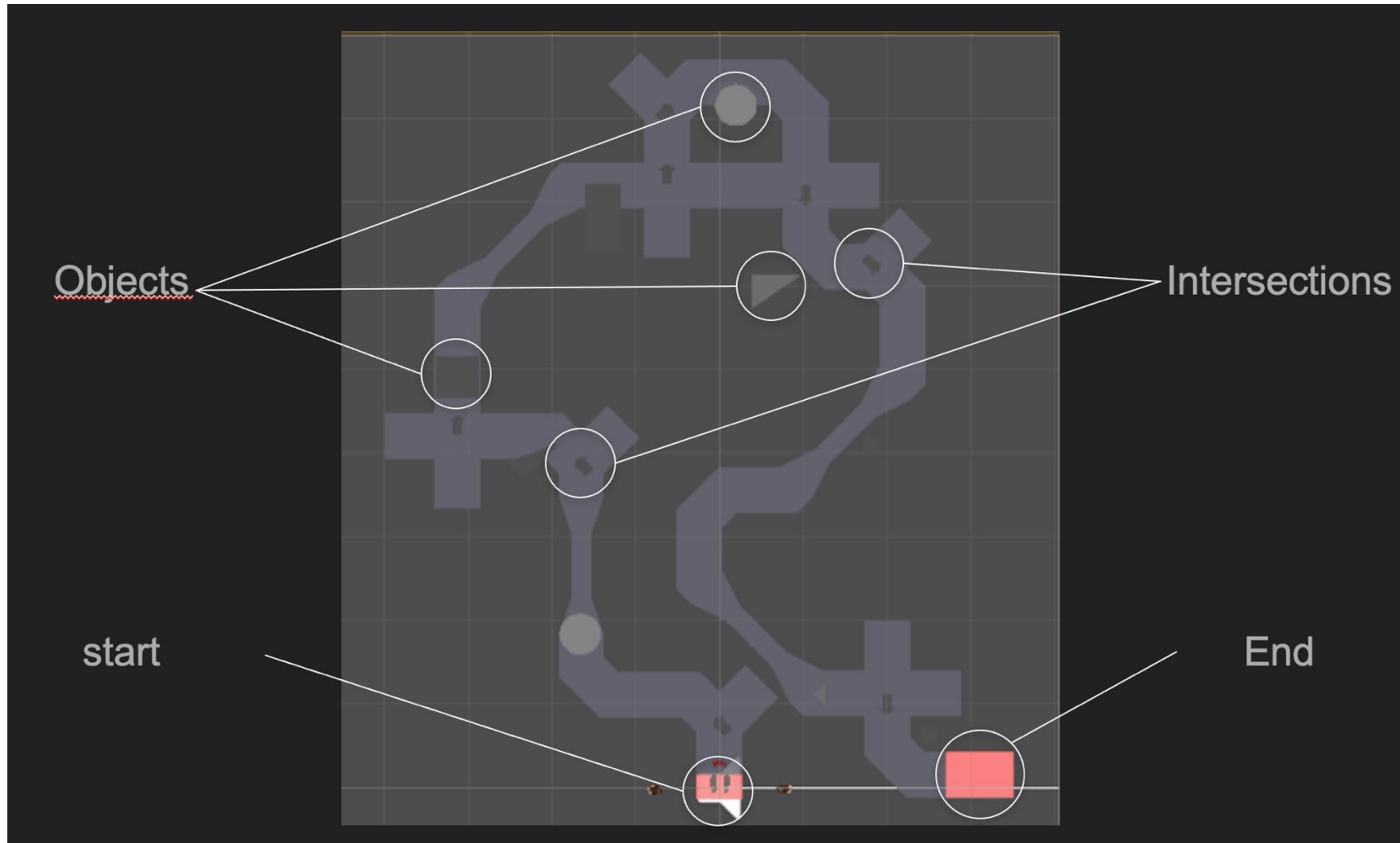
Neuropathie
optique
inflammatoire
anti-MOG

Neuropathie optique
compressive / lymphome
orbitaire bilatéral

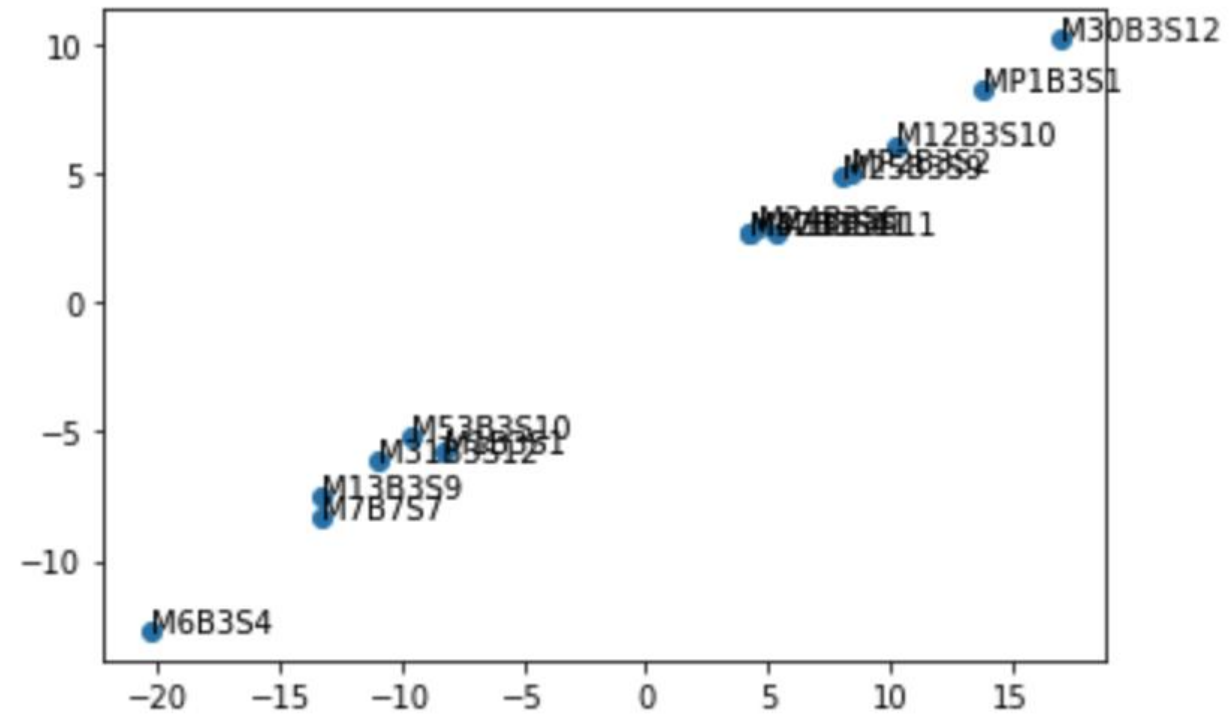
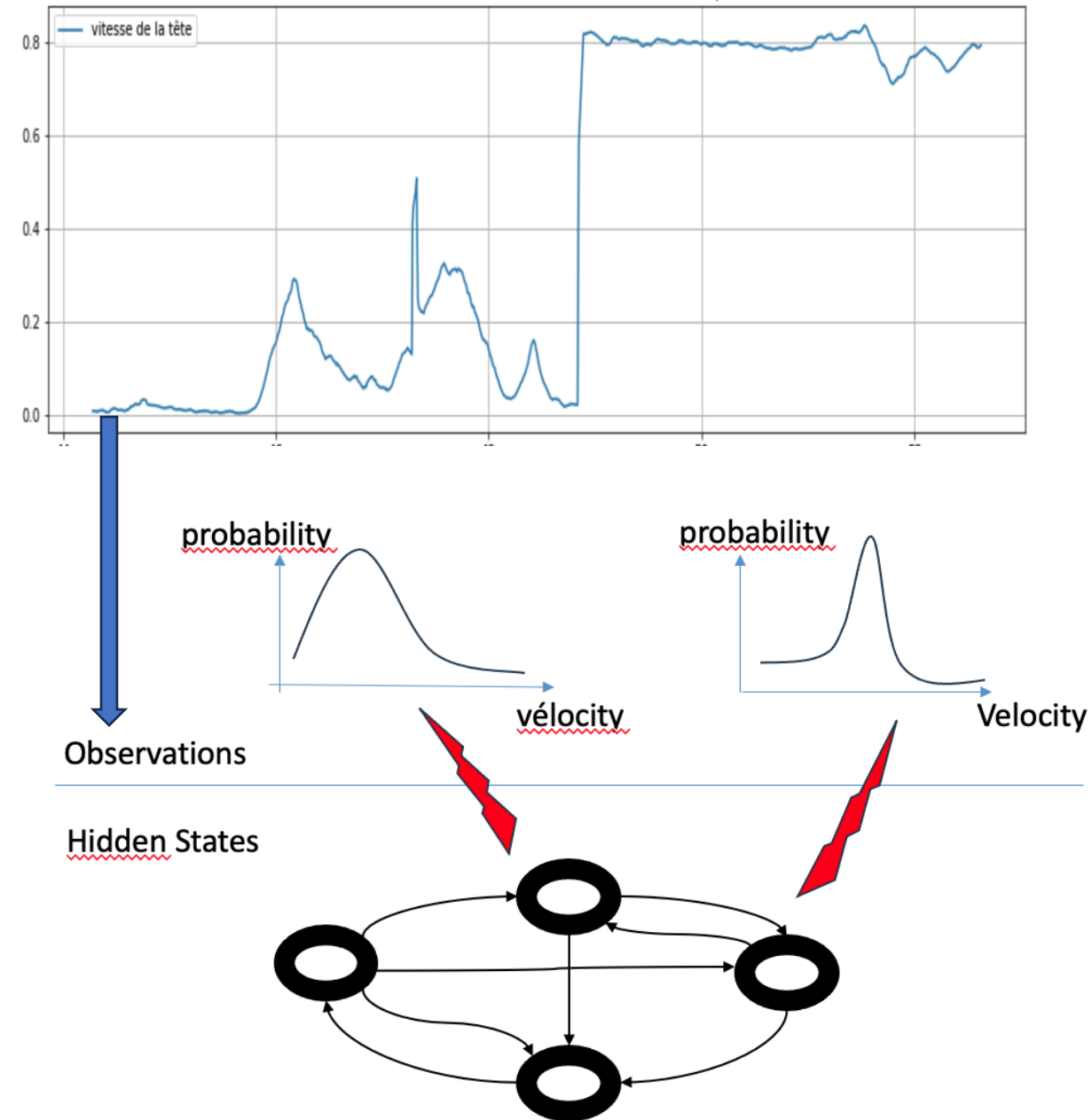
Neuropathie
optique
héréditaire OPA-1



ADVANCED CLUSTERING OF BEHAVIORAL DATA



ADVANCED CLUSTERING OF BEHAVIORAL DATA (HMM+ T-SNE)



Neurosciences vs vision par ordinateur : regard croisé sur l'attention visuelle

Parlons-nous de la même chose?

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

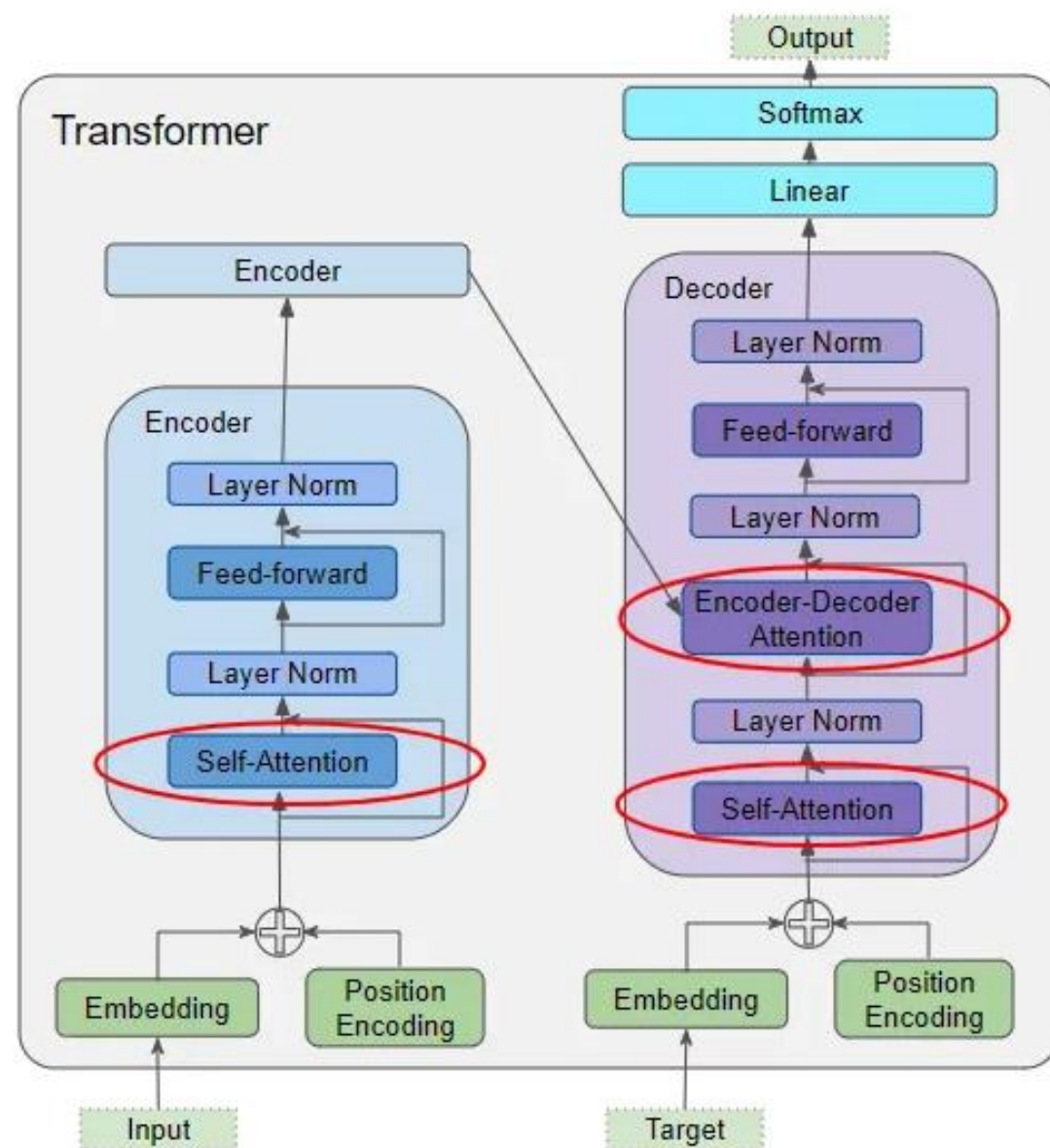
Aidan N. Gomez*[†]
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin*[‡]
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.



Vision Transformers are Inherently Saliency Learners

Yasser Abdelaziz Dahou Djilali^{1,2}
yasser.dahoudjilali2@mail.dcu.ie

Kevin McGuinness¹
kevin.mcguinness@dcu.ie

Noel O'Connor¹
Noel.OConnor@dcu.ie

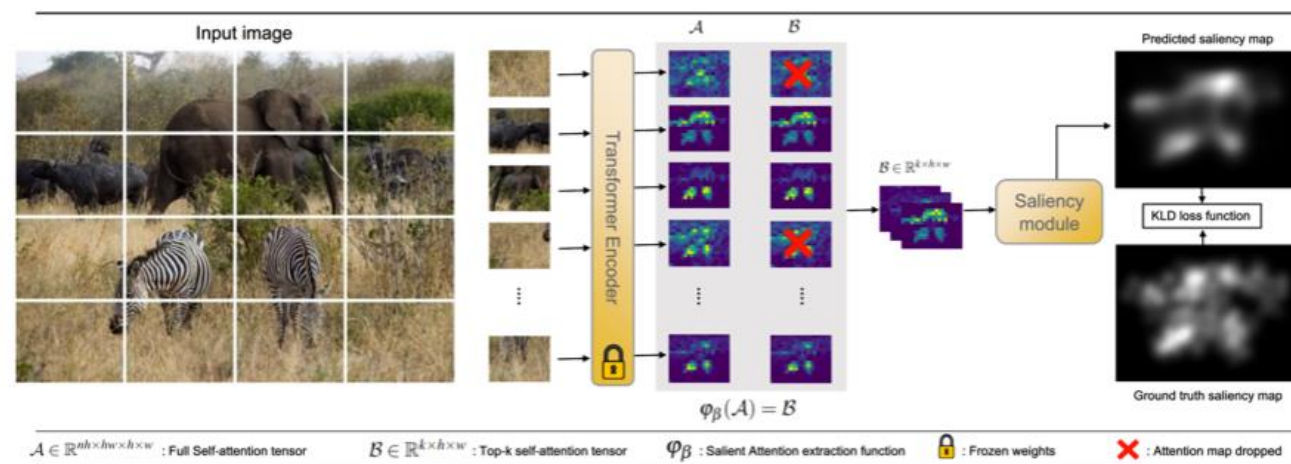
¹ Dublin City University (DCU),
Dublin, Ireland.

² Technology Innovation Insitute,
Abu Dhabi, UAE.

Abstract

Training a Convolutional neural network's (CNNs) auto-encoder has been the defacto approach for visual attention modelling. Recently, (Vision) Transformer models (ViT) achieved excellent performance on various computer vision tasks. In this context, the self-attention mechanism plays a crucial role enabling early aggregation of global information, and ViT residual connections strongly propagate features from lower to higher layers. This raises two important questions: are Vision Transformers inherently learning saliency maps? Are the self-attention maps focusing on the salient regions of the input image? Analyzing the self-attention maps of a pretrained ViTs on saliency prediction datasets, we find that smoothing the internal attention maps with a small number of convolutional filters can achieve reasonable saliency maps with acceptable metric scores. We explore how this phenomenon arises, finding that self-attention promotes early aggregation of global information, then in higher layers, it associates highly attended features, compares their dependencies, and makes analogies over the recurring patterns. This suggests that ViTs first perform feature search, followed by conjunction search combining multiple features sharing higher mutual information. We study the analogies between the self-attention maps and the human generated saliency maps, and conclude with a discussion on the relationship to human visual attention such as feature integration theory.

Figure 3.5: Complete pipeline for training. The input image is processed with a Vision Transformer model to get the self-attention maps tensor $\mathcal{A}$. φ_{β} serves as a medium to produce the Salient Attention maps tesnor $\mathcal{B}$ by selecting top-k informative modes. Then, $\mathcal{B}$ is fed to the Saliency Module to predict the visual saliency map.



Are ViTs inherently saliency learners? (Using Salicon and MIT300 for testing)

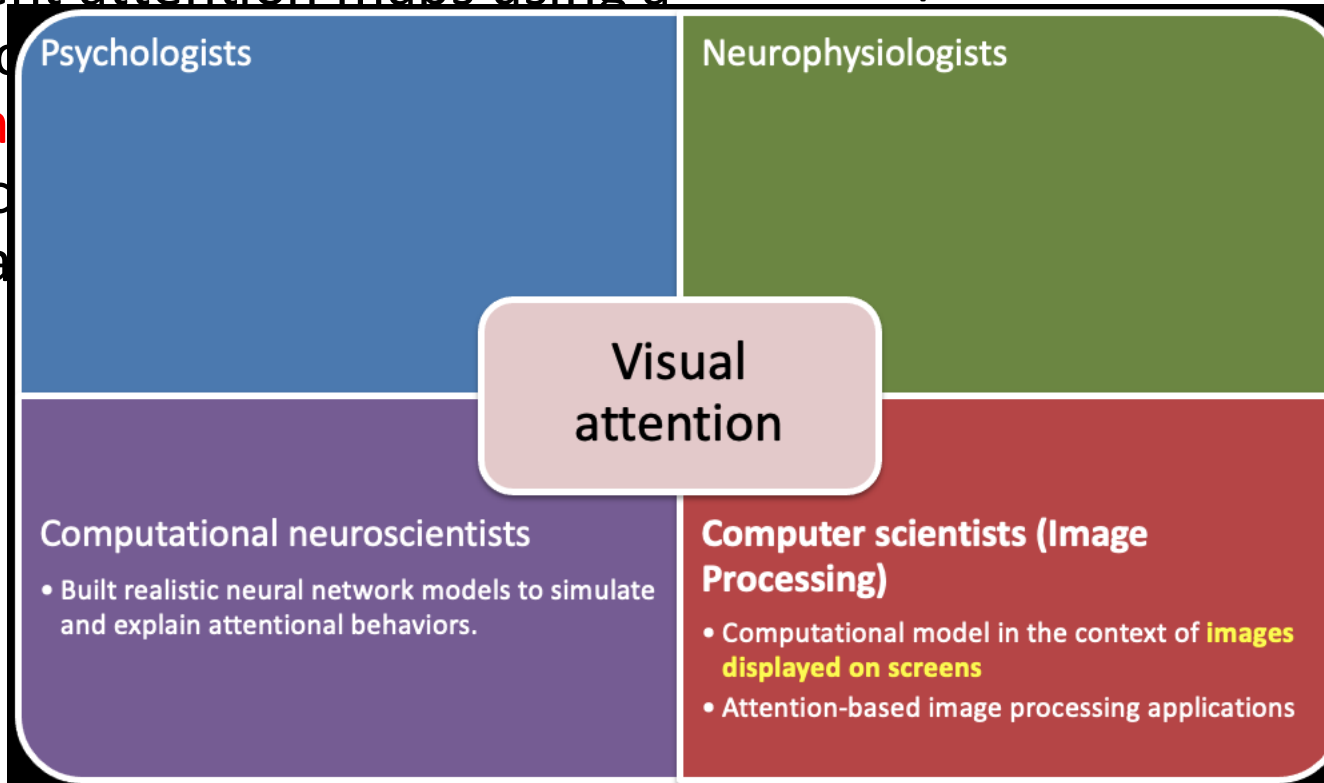
« By analyzing how local/global spatial information is utilised, we show that the the ViT penultimate self-attention maps are similar to the saliency maps »

«retrieving the salient attention maps using a tensor decomposition
tensor of saliency in
mapped to a saliency
convolutional neural

« Our findings suggest that ViTs under a self-supervised regime (DINO self supervised weight) can better capture a wide range of salient features across various contexts,

compared to their supervised

ization with ImageNet
focus on the most
atures linked directly to





The IEEE International Conference on Multimedia & Expo (ICME) is the flagship multimedia conference sponsored by four IEEE societies since 2000, promoting the exchange of the latest advances in multimedia technologies, systems, and applications from both research and development perspectives. ICME 2025 will be held in Nantes, France, from June 30 to July 4, 2025. ✓ will bring together leading researchers and practitioners to share the latest developments and advances in the discipline. Featuring high-quality oral and poster sessions, world-class keynotes, exhibitions, demonstrations, and tutorials, the conference will attract leading researchers and global industry figures, providing excellent networking opportunities. In addition, exceptional papers and contributors will be selected and recognized with prestigious awards.



Ressources rares, cerveaux fatigués, société vieillissante : une IA post-croissance ?

Pr. Patrick Le Callet

Keynote

10 juin 2025



institut
universitaire
de France

