

EduRisk: un pipeline pour simplifier l'accès à la connaissance du risque.

Rapport pour le groupe N03 :



Daniel Nissille



Open Science automne 2025

Sous la supervision du Prof. François Grey

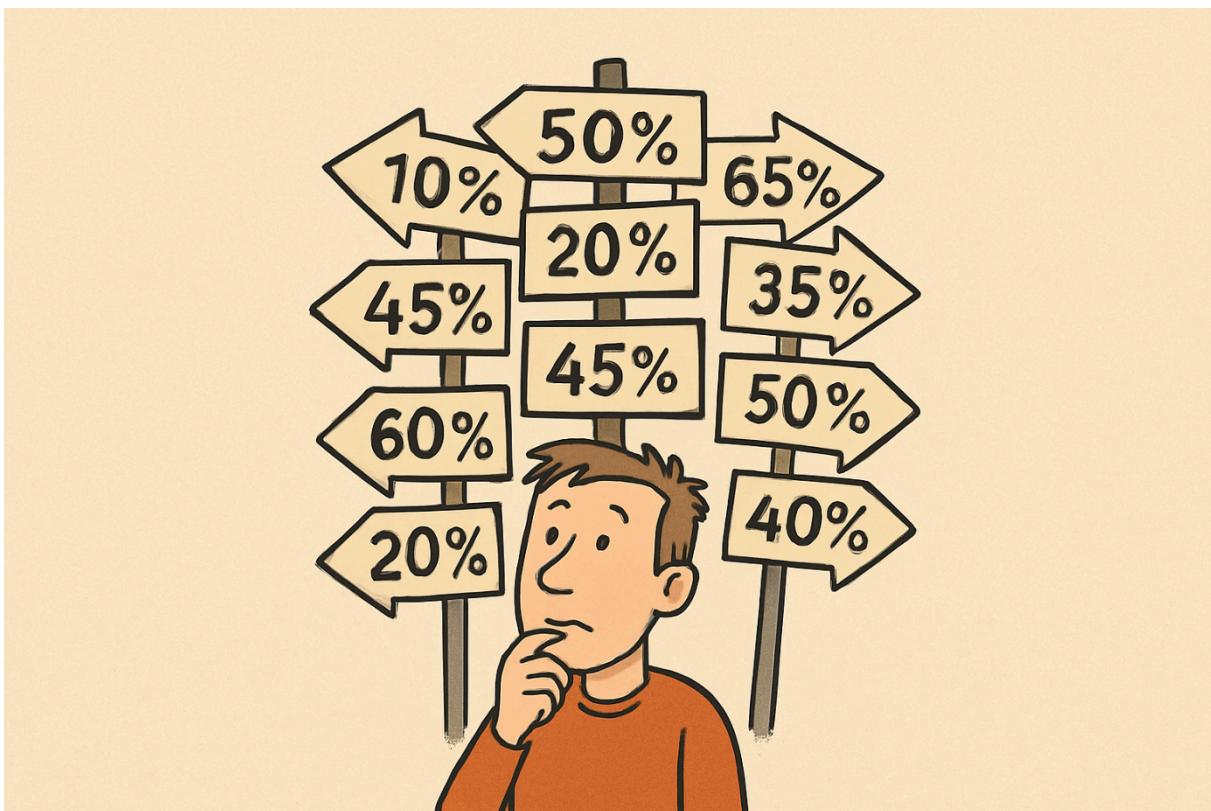
Daniel Nissille, Nam Thang Nguyen, Joël Favre, Mihaela Satcovschi, Loic Kiluangu

Et avec la participation durant le Hackathon de Thibaut De Mallac de Vessac et Julie Hyunh
Khanh-Thu

Résumé du sujet concerné :

Nous vivons à une époque dans laquelle il n'a jamais été autant donné à l'humanité d'être bombardé d'informations. Réseaux sociaux, plateformes de diffusion ou encore journaux, partout des digits, des calculs et des pourcentages. Mais lorsqu'il est question d'aborder des questions de risques les choses se complexifient. Le cerveau humain n'est pas efficace pour traiter l'information probabiliste et même des scientifiques peuvent être biaisés ou égarés par des résultats peu intuitifs surtout s'ils n'ont pas été formés pour ce faire. ([Source des tests de résilience effectués au niveau des uni](#)) [1]. Effectivement si votre magasin favori affiche : « Méga réduction de -40% sur les déjà -40% » vous aurez vite fait de vous dire qu'il s'agit d'une réduction de 80% alors qu'en réalité elle n'est que de 64%. Il en va de même pour des cas concrets de présentation des risques : « Ce vaccin protège à hauteur de 50% » ou encore « Prendre l'avion est beaucoup moins risqué que la voiture. » On comprend ici qu'il faut pouvoir décortiquer l'information pour la rendre plus intuitive, plus accessible dans son ensemble. Combien de personnes sont réellement malades et combien de ces malades peuvent espérer être réellement protégés ? Peut-on comparer de manière intuitive les risques de prendre l'avion avec ceux de prendre la voiture ? La « risk literacy » est justement un domaine qui cherche à mieux définir et rendre les questions liées au risque plus intuitives [2].

Figure 1 Une personne perdue devant la masse d'information du quotidien.



Introduction de la problématique :

La problématique initiale que nous avons choisis de traiter était : “**EduRisk: Cognitive Models for Risk Literacy** AI tools explore how education shapes risk perception and decision-making, using open data to improve teaching on uncertainty and behavioral resilience.”

Il n'a pas été ais  de travailler cette probl matique complexe qui demandait  tre d compos e  en plusieurs points. La “risk literacy” n'est en effet pas juste traductible en fran ais par compr hension du risque. C'est une v ritable science combinant maths, psychologie et esprit critique. [2] L'id e de base ici  tant visiblement de parvenir   valuer l'efficacit  de l'enseignement d'une m thode permettant d'am liorer la r silience devant les questions de risques. Cela pouvant tant avoir un effet sur les d cisions communes du quotidien que sur des d cisions impactant notre parcours de vie. Un point  galement important est l'int gration d'outils IA et de donn es ouvertes.

Pendant le Hackathon nous avons rapidement  t  confront s   un probl me initial. Avant m me de parler de risque, il faut d j  d finir un domaine concern  par ce risque (M decine, conomie...) et avec cela il faut aussi pouvoir trouver des publications associ es. Lors de nos recherches nous avons r alis  que ces dites publications pour autant qu'elles soient accessibles n' taient pas toujours tr s intuitives. Nous avons par cons quent d cid  de nous focaliser sur une premi re  tape fondamentale : l'extraction des donn es li es au risque. Avec l'objectif de

pouvoir enseigner ou clarifier l'information, nous avons imaginé de rendre les données extraites plus intuitives. Nous avons aussi discuté des méthodes qui pourront être mises en place afin de tester l'efficacité du procédé que nous cherchons à développer.

En clair : Nous avons développé l'idée d'une démarche permettant de simplifier l'accès et la compréhension à la connaissance scientifique traitant du risque. Notre pipeline utilise des méthodes de risk literacy adjointes d'un traitement d'extraction par IA. Avec l'intention de pouvoir dans un deuxième temps aussi évaluer l'efficacité de la méthode considérée.

Méthodologie :

1) Idée conceptuelle globale du pipeline :

La discussion avec Madame Kerstin Preuschoff et deux experts en IA nous a conduit à revoir à la baisse la technicité de notre pipeline. Ainsi nous en sommes arrivés à la démarche globale suivante :

- 1) Extraire l'information contenue dans une publication traitant du risque, on parle ici de « scraping IA ».
- 2) Représenter l'information extraite de manière intuitive en utilisant des techniques développées par la « risk literacy ». Par exemple les « facts box » de G.Gigerenzer. [4]
- 3) Tester l'efficacité de la méthode utilisée à l'aide d'un questionnaire et deux groupes tests.
- 4) Ces trois derniers points ayant à être le plus ouverts possible et donc être disponibles gratuitement et avec un maximum de transparence.

L'idée d'appliquer les facts box pour présenter des données médicales n'est pas nouvelle [3] mais l'approche IA pour les construire rend le cas plus user-friendly. Néanmoins, l'idée de notre démarche n'est pas de laisser un chabot effectuer des recherches à notre place, mais plutôt de lui fournir directement les publications scientifiques d'intérêts. Cela ayant pour avantage de ne pas tomber sur des résultats

qui seraient hallucinés puisque c'est visiblement un problème actuel [7] et facilitera aussi les éventuelles améliorations futures de solidification des résultats.

Explication des pipelines et résultats du hackathon.

2) Idée conceptuelle globale de la phase de test :

Est-ce que quelqu'un a trouvé des exemples ?

Notre pipeline de test se voulait très simple : Etablir un questionnaire sur les résultats d'une publication traitant du risque et soumettre ce questionnaire à deux groupes. L'un ayant eu accès à un résultat simplifié par notre pipeline EduRisk et le second sans outil. Les performances de réponses seront alors comparées par exemple avec un test de Student [8]. Afin d'améliorer la mesure de la capacité à prendre des décision d'EduRisk la variable temporel a aussi été considérée.

Results (prototype, mock up) :

Afin de résumer nos attentes voici un petit exemple concret :

Joe est un senior et il aimerait savoir s'il est avantageux de se faire vacciner contre la grippe. Il a lu quelque part que le vaccin protégeait à 60% contre la grippe (influenza) et par conséquent il a l'impression que ce dernier est très efficace, voir même fondamentale. Puisqu'il est un peu sceptique Joe en a parlé à son médecin qui lui a alors recommandé une publication médicale «**Vaccines for preventing influenza in the elderly** » [5] Le problème pour Joe c'est qu'il ne comprend vraiment rien aux statistiques médicales. Avec notre pipeline EduRisk, Joe peut générer une fact box clair qui lui permettra de se faire un avis tranché sur la situation (figure 2).

Figure 2 un exemple de fact box intuitive de la publication de Demicheli et al., générée par ChatGPT.

Fact-box « Si je suis une personne âgée (≥ 65 ans) : vacciner ou non contre la grippe »

Situation / question	Sans vaccin	Avec vaccin	Réduction (absolue / relative)
Population étudiée	—	—	~ 5 000 personnes (8 études) cochrane.org +1
Risque de contracter la grippe sur une saison	~ 6 personnes sur 100	~ 2.4 personnes sur 100	Réduction absolue ≈ 3.6 points % Réduction relative ≈ 60 %
Risque de syndrome grippal (symptômes)	~ 6 personnes sur 100	~ 3.5 personnes sur 100	Réduction absolue ≈ 2.5 points % Réduction relative ≈ 42 %
Pour « 100 personnes âgées vaccinées » → combien évite-t-on un cas?	—	—	~ 1 cas de grippe évité pour 30 vaccinés ~ 1 cas de syndrome grippal évité pour 42 vaccinés

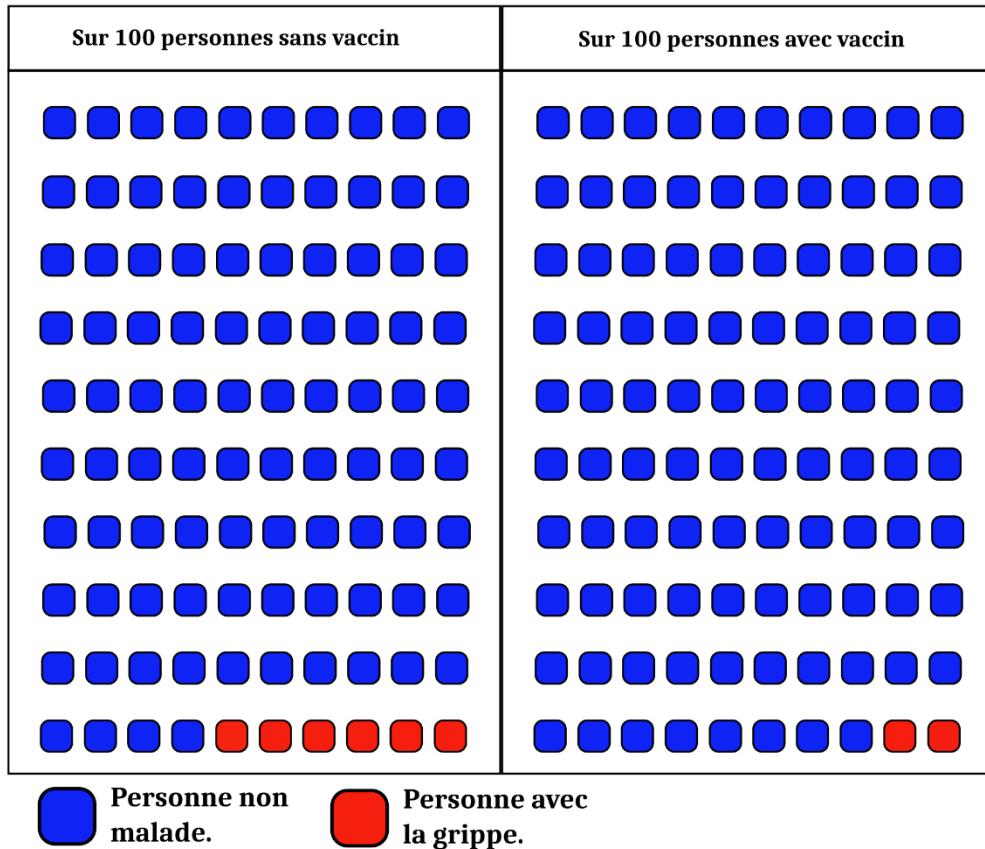
Interprétation simplifiée (en langage courant)

- Si 100 personnes âgées comme vous ne sont **pas vaccinées**, environ 6 vont attraper la grippe en une saison.
- Si ces 100 personnes sont **vaccinées**, environ 2 à 3 d'entre elles attraperont la grippe — soit **3 à 4 gripes évitées**.
- En moyenne, il faut vacciner **30 personnes âgées** pour prévenir **1 infection grippale** en une saison.

Note Figure 2 : La représentation des données sous la forme de fact box permet de montrer les ratios réels et de surmonter une technique souvent utilisée par les industrielles pour venter l'efficacité de leur produit. En effet dans ce cas le fait de se vacciner peut diminuer le nombre de contagion de 4 sur 100 et comme 6 personnes sur 100 risquent de tomber malade il est juste de dire que la réduction relative est de 60%. Mais on réalise tout de suite que cette manière de faire est plus « vendueuse » que de dire : 2 personnes tomberont malade au lieu de 6 sur 100. La simplification en langage courant et la mention du nombre de personnes totales dans l'étude augmente aussi la pertinence de l'analyse considérée. Joe a donc à présent une vision moins biaisée et plus intuitive de la situation, il pourra donc prendre une décision sincère et éclairée.

La représentation en fact box n'est pas la seule manière de procéder, il est aussi possible d'utiliser des représentations en graphiques dit intuitifs parfois plus visuels (Figure 3).

Figure 3 un exemple graphique intuitif.



Note pour la figure 3 : Ce graphique intuitif et simple construit à la main avec INKSCAPE [6] résume encore plus simplement les résultats discutés ci-dessus.

Ici il faut aussi montrer ce qu'on a déjà réussi à faire avec le code. On doit leur vendre du rêve ^^.

Résumé et perspective d'avenir :

Ici faire une petite synthèse de ce qu'on est parvenu à faire et comment on imagine pouvoir faire avancer le projet dans le futur.

- Attention les fact boxes ne sont en rien une garantie que la publication de base est un bon travail, si cette dernière comporte des biais, la méthode employée ici ne les révélera pas.

Je suppose qu'ici on pourrait inclure les conclusions des études qu'on a trouvé(je trouve que ça fait une meilleure division de l'information avec laquelle on a travaillé)-M

Les points 6 et 7 seront traités durant l'écriture du rapport et la bibliographie est en dessous.

Compte rendu du travail par membre :

-Daniel Nissille : A grandement participé à la recherche de publications et d'informations pour mieux cerner la problématique. A œuvré à la compréhension et définition simple de la « risk literacy » et sur le comment transformer la problématique en un objectif atteignable dans les temps impartis. Travail conséquent sur la Rédaction du rapport et de la présentation. Construction du cas exemple de fact boxe sur la grippe et les exemples de référence.

-A vous de jouer :

Bibliographie :

- 1) Cokely, E. T., Feltz, A., Ghazal, S., Allan, J. N., Petrova, D., & Garcia-Retamero, R. (2018). "Skilled Decision Theory: From Intelligence to Numeracy and Expertise." In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 476–505). chapter, Cambridge: Cambridge University Press. [en ligne] url:
<https://www.cambridge.org/core/books/abs/cambridge-handbook-of-expertise-and-expert-performance/skilled-decision-theory-from-intelligence-to-numeracy-and-expertise/C7C6273C841BB197ABF5346F8527217C>
- 2) RiskLiteracy.org : <http://www.riskliteracy.org/>
- 3) Gigerenzer, G., & Kolpatzik, K. (2017). "How new fact boxes are explaining medical risk to millions." BMJ (Clinical research ed.), 357, j2460.
<https://doi.org/10.1136/bmj.j2460>
- 4) How to do facts box:
<https://www.hardingcenter.de/en/transfer-and-impact/fact-boxes>
- 5) Demicheli V, Jefferson T, Di Pietrantonj C, Ferroni E, Thorning S, Thomas RE, Rivetti A. «Vaccines for preventing influenza in the elderly. » Cochrane Database of Systematic Reviews 2018, Issue 2. Art. No.: CD004876. DOI: 10.1002/14651858.CD004876.pub4. Accessed 07 December 2025.
- 6) Inkscape Project. (2025). *Inkscape*. Retrieved from <https://inkscape.org>
- 7) Ayush Agrawal, Mirac Suzgun, Lester Mackey, and Adam Kalai. 2024. "Do Language Models Know When They're Hallucinating References?." In *Findings of the Association for Computational Linguistics: EACL 2024*, pages

- 912–928, St. Julian's, Malta. Association for Computational Linguistics.
Accessed 21 December 2025
- 8) Test de Student:https://fr.wikipedia.org/wiki/Test_de_Student, Accessed 21 December 2025

Figures:

Les figures 1 et 2 ont été générée avec ChatGPT : OpenAI. (2025). *ChatGPT* [Grand modèle linguistique]. Récupéré en décembre 2025, de <https://chat.openai.com/>.

La figure 3 a été construite à la main avec INKSCAPE voir référence [6]