Dossier de Conception :

Interaction Humain-IA pour l'Application de Suivi Nutritionnel

NutriSnap-HAI

Soumis par :

Shirel AMOZIEG
Sama SATARIYAN
Erisa KOHANSAL

Sous la direction de :

François Bouchet

Sommaire

1.0 Introduction et Objectif du Projet

Ce projet, nommé ***NutriSnap-HAI***, s'inscrit dans le champ de recherche de l'Interaction Humain-IA (HAI) et a pour but de concevoir et d'évaluer une application de suivi nutritionnel. L'objectif principal de cette étude n'est pas d'évaluer la performance technique d'un modèle d'intelligence artificielle, mais d'analyser en profondeur les dynamiques d'interaction entre un utilisateur et un système d'IA dans le cadre d'une tâche d'estimation nutritionnelle à partir d'images. Nous cherchons à comprendre comment concevoir une collaboration efficace, où l'IA assiste l'utilisateur sans le déposséder de son contrôle et de sa capacité de décision.

La problématique centrale de cette étude se décline en plusieurs questions de recherche clés, qui ont guidé la conception du prototype et le protocole expérimental :

- **Analyser les conditions** dans lesquelles l'utilisateur choisit de suivre, de corriger ou de rejeter les suggestions de l'IA.
- **Évaluer comment l'affichage explicite de l'incertitude** de l'IA influence le comportement et la prise de décision de l'utilisateur.
- **Mesurer la "Reliance"** comme métrique principale pour quantifier la confiance appropriée de l'utilisateur envers le système.
- **Valider l'efficacité des mécanismes de contrôle** (garde-fous) conçus pour garantir que la décision finale reste entre les mains de l'utilisateur.

Cette section introductive a posé les fondations de notre démarche. Nous allons maintenant présenter la logique de conception et d'implémentation du prototype expérimental qui a permis d'explorer ces questions.

2.0 Logique de Conception et d'Implémentation

La conception d'une interaction saine entre un humain et une IA repose sur des choix stratégiques qui visent à favoriser la confiance, la transparence et le contrôle. Pour le prototype *NutriSnap-HAI*, chaque décision de conception a été guidée par les principes de l'Interaction Humain-IA, avec l'ambition de créer une expérience où l'IA n'est pas une boîte noire, mais un partenaire compréhensible et maîtrisable. Ces éléments de conception : le contrat d'interaction à trois boutons et le badge d'incertitude, ont été conçus comme des leviers expérimentaux, permettant d’observer et de mesurer la reliance de l’utilisateur au système, ainsi que ses réactions face à des situations d’incertitude.

2.1 Le Périmètre Testé : L'Identification Rapide du Repas

Pour garantir une analyse rigoureuse, l'étude s'est concentrée sur une fonctionnalité précise et critique de l'application : l'identification semi-automatisée d'un repas à partir d'une photographie. Le but n'était pas de développer une application de suivi nutritionnel complète, mais de créer un prototype expérimental focalisé sur ce point d'interaction. Tandis que la vision globale du projet inclut une approche de suivi nutritionnel personnalisée et potentiellement médicalisée, le périmètre de ce prototype a été délibérément restreint à l'estimation du nom du plat, des calories et des macronutriments (protéines, glucides, lipides). Cette approche nous a permis d'isoler les variables d'interaction et de collecter des données ciblées sur le comportement de l'utilisateur face aux suggestions de l'IA.

2.2 Le Contrat d'Interaction Humain-IA

Le flux d'interaction a été défini de manière claire et structurée pour l'utilisateur, en trois étapes séquentielles :

1. **Saisie (Input) :** L'utilisateur prend une photographie de son plat. Cette action constitue le déclencheur explicite de l'interaction.
2. **Suggestion de l'IA :** Le système analyse l'image et propose une estimation complète (nom du plat, calories, macronutriments). Crucialement, cette suggestion est accompagnée d'un badge d'incertitude actionnable ("Low", "Medium", ou "High") qui communique le niveau de confiance de l'IA.
3. **Décision de l'Utilisateur :** Face à cette suggestion, l'utilisateur dispose de trois actions claires et distinctes pour exercer son contrôle :
    - **"OK" :** Accepter l'estimation de l'IA dans son intégralité.
    - **"ALMOST THERE" :** Accepter la base de l'estimation mais souhaiter l'ajuster ou la corriger (override).
    - **"NO" :** Rejeter complètement la suggestion de l'IA.

2.3 Le Choix du "Magicien d'Oz"

Pour simuler le comportement de l'IA, nous avons opté pour la méthode du "Magicien d'Oz" (Wizard of Oz - WoZ). Dans cette approche, un expérimentateur humain (le "Magicien") génère en temps réel les réponses de l'IA en se basant sur la photo soumise par l'utilisateur. Ce choix était méthodologiquement crucial car il nous a permis de dissocier l’analyse de l’interaction utilisateur-IA de la performance d’un modèle d’apprentissage automatique particulier, et ainsi de considérer les dynamiques

d’interaction comme la variable principale de l’étude. Cette méthode offre un contrôle total sur les scénarios expérimentaux, permettant de simuler délibérément des erreurs ou de faire varier les niveaux d'incertitude pour observer les réactions des utilisateurs dans des conditions maîtrisées.

Ces choix de conception ont permis de créer un environnement expérimental robuste, servant de base à un protocole rigoureux de collecte de données.

3.0 Protocole Expérimental

Pour collecter des données objectives sur le comportement des utilisateurs, nous avons mis en place un protocole expérimental structuré. Ce protocole a été spécifiquement conçu pour comparer la performance et les comportements des utilisateurs lorsqu'ils interagissent avec le système, avec ou sans l'assistance de l'intelligence artificielle.

3.1 Conditions Expérimentales

L'étude a été menée en comparant deux groupes distincts pour isoler l'impact de l'IA :

- **Groupe 1 (Condition Humain (H_only) - Contrôle) :** Dans cette condition, l'utilisateur doit identifier et saisir manuellement les informations de son repas après avoir pris une photo, sans recevoir aucune suggestion de l'IA. Ce groupe sert de baseline pour mesurer la performance d'une tâche de saisie manuelle.
- **Groupe 2 (Condition IA (H+IA) - Expérimental) :** L'utilisateur reçoit une suggestion de l'IA (simulée par le Magicien d'Oz) après avoir pris sa photo. Sa tâche consiste alors à évaluer cette suggestion et à choisir l'une des trois actions proposées (accepter, ajuster, rejeter).

3.2 Déroulement d'un Essai

La tâche demandée à chaque participant était simple : prendre en photo un plat fourni par les expérimentateurs et l'enregistrer dans l'application le plus efficacement et précisément possible. L'expérience s'est déroulée à l'aide d'un prototype fonctionnel développé avec la technologie *Streamlit*. En arrière-plan, l'expérimentateur jouant le rôle du Magicien contrôlait les réponses de l'IA via une interface dédiée (la barre latérale de l'application).

3.3 Rôle du Magicien d'Oz (WoZ)

Depuis son interface de contrôle, le Magicien effectuait plusieurs actions clés pour simuler une IA crédible et pour enregistrer des données essentielles à l'analyse :

- Il analysait la photo prise par l'utilisateur.
- Il définissait la sortie de l'IA : le texte descriptif du plat, les calories et les macronutriments.
- Il choisissait le niveau d'incertitude à afficher à l'utilisateur (Low, Medium, High).
- Il indiquait si sa propre estimation (celle de l'IA simulée) était objectivement correcte (Y) ou incorrecte (N), une information cruciale et invisible pour l'utilisateur, mais indispensable pour le calcul ultérieur de la métrique de Reliance.

Ce protocole nous a permis de recueillir un ensemble de données précises et structurées, que nous détaillons dans la section suivante.

4.0 Format et Volume des Données Collectées

Chaque interaction effectuée par les utilisateurs avec le prototype NutriSnap-HAI a été systématiquement enregistrée dans le but de permettre une analyse quantitative rigoureuse et reproductible. Toutes les données ont été consolidées dans un unique fichier au format .csv, nommé ***logs.csv***.

Au total, nous avons collecté un volume de **79 interactions** distinctes, réparties sur **7 sessions** utilisateur. Chaque ligne du fichier de logs représente une décision finale prise par l'utilisateur pour un essai donné.

Le tableau ci-dessous présente les champs de données les plus importants qui ont été collectés pour chaque interaction, accompagnés d'une brève description.

Champ	Description
session_id	Identifiant unique et anonyme pour chaque session utilisateur.
condition	Condition expérimentale (IA (H+IA) ou Humain (H_only)).
ai_uncertainty	Niveau d'incertitude de l'IA affiché à l'utilisateur (Low, Medium, High).
correct	Indique si la proposition de l'IA était correcte (Y) ou non (N).

human_action	L'action choisie par l'utilisateur (accept, override, reject).
decision_time_ms	Temps écoulé (en millisecondes) entre l'affichage de la suggestion et la décision de l'utilisateur.
human_intervention	Indicateur binaire (1 ou 0) signalant une saisie manuelle.

Ces données structurées constituent la base sur laquelle nous avons pu tester un ensemble d'hypothèses spécifiques sur l'interaction Humain-IA.

5.0 Hypothèses Testées

L'analyse des données collectées a été guidée by un ensemble d'hypothèses pré-définies, visant à évaluer l'efficacité de l'assistance de l'IA et la qualité de l'interaction conçue. La validation ou l'invalidation de ces hypothèses, dont le détail technique se trouve dans le notebook d'analyse joint à ce rapport, permet de tirer des conclusions concrètes sur la conception du système.

1. **Hypothèse 1 (Performance) :** L'assistance de l'IA (H+IA) permet de réduire significativement le temps nécessaire pour enregistrer un repas par rapport à une saisie entièrement manuelle (H_only).
2. **Hypothèse 2 (Confiance Appropriée) :** Les utilisateurs feront preuve d'une "Reliance" positive, c'est-à-dire qu'ils accepteront plus fréquemment les suggestions de l'IA lorsque celles-ci sont correctes que lorsqu'elles sont incorrectes.
3. **Hypothèse 3 (Influence de l'Incertitude) :** Le niveau d'incertitude communiqué par l'IA influence le comportement de l'utilisateur. Une incertitude plus élevée entraînera une augmentation du temps de décision et un taux de correction (override) plus important.
4. **Hypothèse 4 (Validation des Garde-fous) :** Les mécanismes de sécurité (garde-fous) implémentés, tels que l'abstention en cas de forte incertitude ou le passage en mode manuel après des échecs répétés, fonctionnent comme prévu et forcent une reprise de contrôle par l'utilisateur.

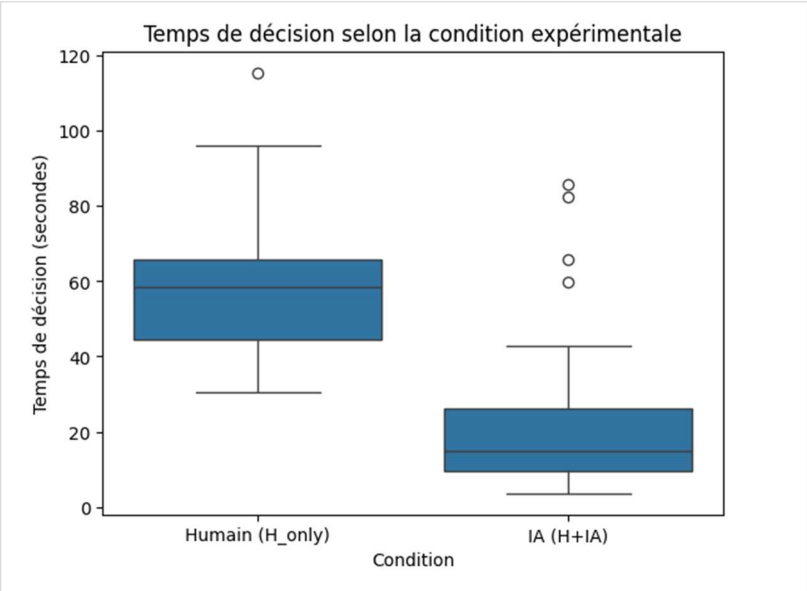
La section suivante présente les résultats détaillés de l'analyse menée pour tester chacune de ces hypothèses.

6.0 Résultats et Discussion

Cette section présente les principaux résultats quantitatifs issus de l'analyse des 79 interactions enregistrées. Chaque résultat est accompagné d'une discussion sur ses implications pour la conception d'interactions Humain-IA efficaces et responsables.

6.1 Analyse de la Performance : L'IA comme Accélérateur

L'analyse de l'impact de l'IA sur le temps de tâche confirme notre première hypothèse. La conclusion principale est sans équivoque : la condition IA (H+IA) est nettement plus rapide que la condition de contrôle Humain (H_only).



Le graphique ci-dessus illustre cette différence de manière frappante. L'IA ne se contente pas de réduire le temps médian ; elle diminue également la variabilité des temps de décision. Ce résultat suggère que l'assistance de l'IA transforme une tâche cognitivement coûteuse de saisie complète en une tâche de vérification beaucoup plus rapide et homogène, où l'utilisateur passe moins de temps à chercher et à taper, et plus de temps à valider une proposition pré-établie.

6.2 Analyse de la Confiance : Une Reliance Modérée

La métrique de "Reliance" mesure la capacité de l'utilisateur à faire confiance à l'IA de manière appropriée. Elle est calculée comme la différence entre la probabilité de suivre l'IA quand elle a raison et la probabilité de la suivre quand elle a tort.

$$Reliance = P(suivre\ IA \mid IA\ correcte) - P(suivre\ IA \mid IA\ incorrecte)$$

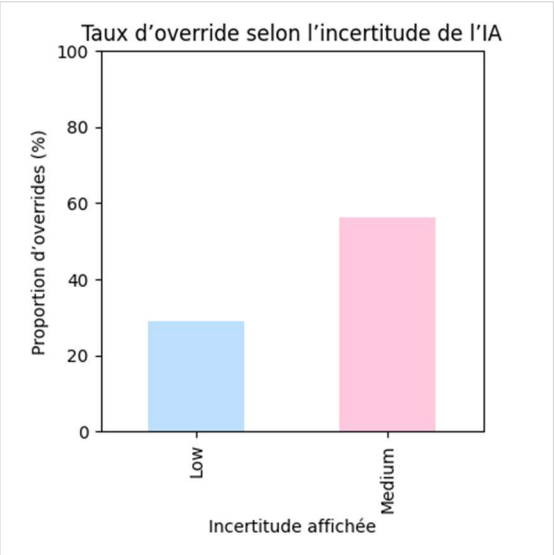
Les résultats numériques clés issus de notre analyse sont les suivants :

- P(suivre IA | IA correcte) : **54 %**
- P(suivre IA | IA incorrecte) : **26 %**
- **Score de Reliance : 0.28**

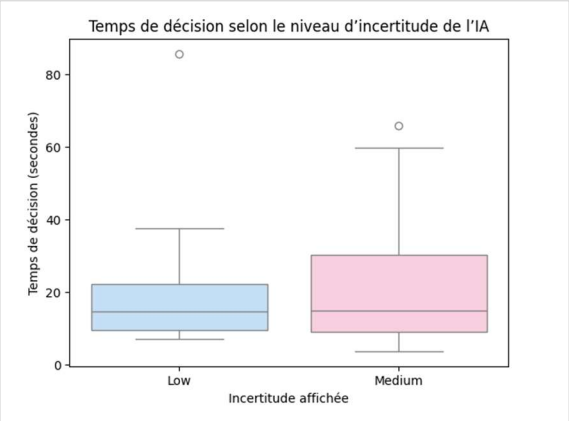
Un score de 0.28 est le signe d'une confiance positive mais modérée. Ce résultat est particulièrement encourageant : il montre que les utilisateurs ne suivent pas l'IA aveuglément. Ils exercent un jugement critique, acceptant plus volontiers les suggestions correctes. Cependant, le fait qu'ils acceptent tout de même une suggestion incorrecte dans 26 % des cas montre que cette confiance n'est pas parfaite. Ce constat souligne l'importance cruciale des fonctionnalités override et reject comme garde-fous contre les biais d'automatisation, garantissant que l'autonomie et le discernement de l'utilisateur sont préservés.

6.3 Comportement face à l'Incertitude

Cette partie explore comment la communication explicite de l'incertitude par l'IA module le comportement de l'utilisateur, validant ainsi notre troisième hypothèse. Le taux d'ajustement (override) passe de **29.03 %** pour une incertitude "Low" à **56.25 %** pour une incertitude "Medium".



Cette augmentation significative prouve que les utilisateurs intègrent le signal d'incertitude dans leur processus de décision. Lorsque l'IA exprime des doutes, les utilisateurs deviennent plus critiques, plus prudents, et sont plus enclins à reprendre le contrôle pour corriger l'estimation.



De plus, le temps de décision médian et sa variabilité augmentent également lorsque l'incertitude passe de "Low" à "Medium", comme le montre le boxplot ci-dessus. Cela confirme que l'utilisateur consacre une charge cognitive plus importante et prend plus de temps pour évaluer la situation lorsque le système lui-même signale une potentielle faiblesse.

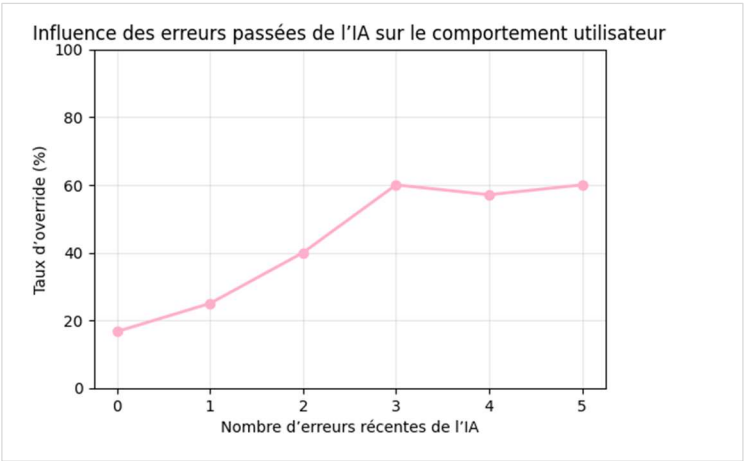
6.4 Analyse Qualitative de l'Action "Override" (Ajuster)

L'action "ALMOST THERE" (override) n'est pas un simple rejet ; elle représente un mécanisme clé de collaboration et de correction fine. L'analyse des 20 cas d'override révèle ce que les utilisateurs choisissent de corriger :

- **Identification du plat : 70 %**
- Calories : 25 %
- Macros (Protéines/Glucides/Lipides) : 15 %

Ce résultat est très instructif : les utilisateurs se sentent majoritairement compétents pour corriger l'identification sémantique du plat (ex: "Pâtes au pesto" au lieu de "Pâtes au fromage"). Cependant, ils ont tendance à faire confiance à l'IA pour les estimations numériques complexes (calories, macros), même lorsqu'ils ajustent l'entrée.

Une analyse plus fine révèle un phénomène inattendu : dans **25 % des cas**, l'action override n'a mené à aucune correction effective des données. Ce comportement suggère que l'override est aussi utilisé comme un mécanisme de vérification et de réassurance. L'utilisateur reprend temporairement le contrôle pour s'assurer de l'exactitude de la suggestion avant de la re-confirmer, illustrant une forme de confiance pragmatique qui va au-delà d'un simple "oui/non".

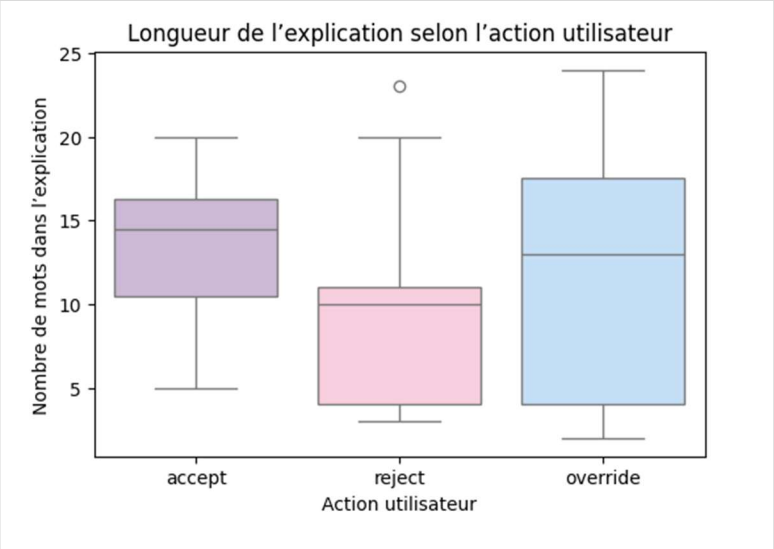


Enfin, l'analyse montre que la confiance est dynamique. Le taux d'override augmente avec le nombre d'erreurs récentes commises par l'IA, prouvant que les utilisateurs adaptent leur niveau de vigilance en fonction de l'historique de performance du système.

6.5 Rôle de l'explicabilité

L'explicabilité, telle que définie par le guiddeline G4, vise à fournir à l'utilisateur une justification utile de la prédiction de l'IA pour éviter l'effet "boîte noire". Dans le prototype NutriSnap-HAI, cette fonctionnalité est matérialisée par la section *Principaux facteurs*, où le système (WoZ) liste les facteurs (ingrédients) ayant mené au calcul des nutriments.

L'analyse des données révèle que la richesse de l'explication fournie par le Magicien (WoZ) influence directement la décision finale de l'utilisateur.



• **Acceptation (OK) :** Les cas d'acceptation sont associés aux explications les plus longues et détaillées, avec une médiane de 14 à 15 mots. Cela suggère que lorsque l'IA justifie sa décision de manière précise, elle renforce la confiance de l'utilisateur et valide son raisonnement.

• **Rejet (NO) :** À l'inverse, les rejets systématiques coïncident avec les explications les plus courtes (médiane de 10 mots). Une justification trop pauvre ou vague est souvent perçue comme une rupture du contrat d'interaction, l'utilisateur ne parvenant pas à valider la prédiction.

• **Ajustement (ALMOST THERE) :** Cette action d'override présente une médiane intermédiaire (12 à 13 mots) mais avec une très forte variabilité. Cela indique une "zone grise" où l'utilisateur comprend partiellement la proposition mais juge nécessaire de reprendre la main pour affiner le résultat.

Ces résultats démontrent que l'explicabilité n'est pas qu'une simple information textuelle, mais un véritable mécanisme de régulation de la confiance. Une explication détaillée favorise une reliance saine, permettant à l'humain de confirmer que les ingrédients détectés correspondent bien à son assiette. Le fait que les utilisateurs acceptent moins les explications courtes prouve qu'ils conservent un esprit critique et ne tombent pas dans une confiance aveugle envers le système.

6.6 Validation des Garde-fous

Les garde-fous sont des filets de sécurité essentiels conçus pour garantir que le contrôle humain est toujours maintenu dans les situations critiques. L'analyse des logs confirme leur efficacité totale, validant notre quatrième hypothèse.

- **GF1 (Échecs Répétés) :** Dans **100% des cas** où deux rejets (NO) consécutifs ont eu lieu, le système a correctement forcé une intervention manuelle lors de l'essai suivant. La boucle d'échecs automatisés a été systématiquement interrompue.
- **GF2 (Incertitude Élevée) :** Dans **100% des cas** où l'incertitude simulée était "High", l'IA s'est abstenue de fournir une estimation. Le système a bloqué toute possibilité d'acceptation automatique et a forcé un passage en mode manuel.

Ces résultats confirment que les mécanismes de sécurité implémentés sont robustes et fonctionnent comme prévu.

7.0 Conclusion et Perspectives

Cette étude a démontré que la conception d'une interaction transparente, où l'IA communique ses limites et où l'utilisateur dispose de moyens de contrôle nuancés, est fondamentale pour une collaboration Humain-IA efficace. L'action override (ALMOST THERE) se révèle être bien plus qu'un simple bouton : c'est un mécanisme d'interaction riche qui incarne le compromis entre automatisation et contrôle. Les résultats de ce travail peuvent être synthétisés en trois principes de conception clés pour les systèmes d'IA collaboratifs :

1. **Communiquer l'Incertitude pour Calibrer la Confiance :** Signaler explicitement le niveau de confiance de l'IA encourage un examen critique de la part de l'utilisateur, prévenant ainsi une confiance aveugle et favorisant une prise de décision plus éclairée.
2. **Fournir un Contrôle Nuancé :** Offrir des actions intermédiaires comme l'override facilite la correction fine et la collaboration, dépassant les modèles binaires d'acceptation/rejet et reconnaissant la complexité de la confiance humaine.

3. **Implémenter des Garde-fous Non Négociables :** Mettre en place des règles strictes (par ex., l'abstention en cas de forte incertitude ou d'échecs répétés) garantit que la supervision humaine est appliquée dans les scénarios où l'automatisation présente un risque.

7.1 Limites du Travail

Il est important de reconnaître les limites de cette étude afin de contextualiser la portée de nos conclusions :

- **Taille de l'échantillon :** L'étude a été menée sur un nombre limité de sessions (7), ce qui ne permet pas de généraliser statistiquement les résultats à une population plus large.
- **Simulation de l'IA :** L'utilisation de la méthode du "Magicien d'Oz", bien que méthodologiquement justifiée, ne reflète pas les imperfections, les biais ou les temps de latence d'un véritable modèle d'intelligence artificielle.
- **Contexte du laboratoire :** L'expérience s'est déroulée dans un environnement contrôlé. Le comportement des utilisateurs pourrait différer dans un contexte d'usage réel (ex: au restaurant, avec une mauvaise luminosité, sous pression temporelle).

7.2 Améliorations et Pistes Futures

Sur la base des enseignements et des limites de ce travail, plusieurs pistes de recherche futures peuvent être envisagées :

- Mener une étude à plus grande échelle avec un nombre de participants plus conséquent pour renforcer la validité statistique des résultats.
- Intégrer un véritable modèle d'IA d'analyse d'images pour étudier les interactions avec une IA non simulée et observer l'émergence de nouveaux comportements face à ses erreurs et biais spécifiques.
- Explorer des mécanismes d'explication (Garde-fou G4) plus riches et interactifs, et évaluer leur impact sur la confiance, la compréhension et la capacité de l'utilisateur à détecter les erreurs de l'IA.
- Déployer le prototype dans des conditions d’usage réelles (ex: suivi longitudinal sur plusieurs jours) pour observer comment la confiance et les habitudes d'interaction évoluent dans un contexte réel.

Enfin, ces travaux pourraient constituer une base pour l'intégration progressive de fonctionnalités de suivi nutritionnel plus personnalisées, tout en conservant les principes de transparence et de contrôle utilisateur mis en évidence dans cette étude.