



REDS MÉTHODOLOGIE RECHERCHE

REDS - introduction

Thursday 19th September, 2024

Laure Soulier



Informations pratiques

Organisation et enseignants

- Méthodologie recherche
 - Enseignant : Laure Soulier (laure.soulier@isir.upmc.fr)
 - Contrôle continu : Soutenance de projet
- Déploiement et MLOps
 - Enseignant : Christophe Boudier (christophe.boudier@isir.upmc.fr)
 - Pas de contrôle continu
- Méthodologie projet
 - Enseignant : Olivier Schwander (olivier.schwander@isir.upmc.fr)
 - Contrôle continu : Soutenance de projet

- Contrôle continu (50%)
 - Soutenance partie recherche (25%)
 - Soutenance parties projet (25%)
- Evaluation finale (50%)

Motivations

- Pourquoi ?
 - Vous faire sortir la tête du guidon
 - Vous apprendre à anticiper sur votre travail scientifique
 - Développer votre créativité
 - Vous sensibilisez à la rigueur scientifique
 - Capitaliser de nouvelles connaissances en machine learning (domaines d'application)
- Pour qui ?
 - Futurs doctorants/chercheurs
 - Futurs salariés du secteur privé
 - Toute personne qui veut s'intéresser à un problème de sciences (dures)

La frontière recherche-industrie est fine

- Une collaboration étroite entre industrie et laboratoires de recherche (thèses CIFRE, projets de recherche, séminaires, ...)
- On trouve des services R&D dans de nombreuses entreprises.
- Il est important de se tenir au courant des avancées du domaine, surtout en machine learning !

Organisation Partie Recherche

- Cours orienté projet/mini-stage (travail en binôme)
 - Méthodologie sur la démarche de recherche scientifique (en informatique, en machine learning)
 - Travail personnel en cours et/ou à la maison
 - Retour global et/ou individuel de l'enseignant sur les différentes étapes du projet
- Evaluation
 - Dernières séances : restitution du travail à l'oral

Réserves

- Ces propos n'engagent que moi !
 - Votre encadrant de stage/thèse/laboratoire peut avoir sa propre méthodologie.
 - Vos futurs collègues en industrie aussi.
 - L'idée est de vous donner un point de départ et vous ouvrir l'esprit.
- Mon discours peut être biaisé
 - Académique, enseignant chercheur en informatique
 - Issue du domaine de la recherche d'information

Qu'est-ce que la recherche ?

Qu'est-ce que la recherche ? [1] – *Définition*

Qu'est-ce que la recherche ? [1] – *Définition*

Le travail de recherche est la construction d'un "objet scientifique":

- Explorer/décrire/expliquer un phénomène
- Résoudre un problème
- Questionner ou réfuter des résultats fournis dans des travaux antérieurs ou une thèse
- Expérimenter un nouveau procédé, une nouvelle solution, une nouvelle théorie
- Appliquer une pratique à un phénomène

La recherche scientifique est un processus dynamique ou une démarche rationnelle qui permet d'examiner des phénomènes, des problèmes à résoudre, et d'obtenir des réponses précises à partir d'investigations. Ce processus se caractérise par le fait qu'il est systématique et rigoureux et conduit à l'acquisition de nouvelles connaissances. Les fonctions de la recherche sont de décrire, d'expliquer, de comprendre, de contrôler, de prédire des faits.

Recherche à plusieurs critères

- Différents niveaux de la recherche : description, classification, explication/compréhension, proposition de modèles
- Modes d'investigation : quali/quant/mixte avec études empiriques, exploratoires, expérimentales, études qualitatives
- Stratégies de vérification : observation, expérimentation, enquête, étude de cas

Qu'est-ce que la recherche ? [1] – *Courants de pensées*

- Claude Bernard (1813-1878)
 - Observation de la réalité est possible sans présupposés ;
 - Formulation d'une hypothèse (théorie) par la créativité du savant ;
 - Vérification expérimentale par confrontation de l'hypothèse avec la réalité (qui a toujours raison).

Inductivisme (raisonnement qui va du particulier au général)

"La meilleure théorie est celle qui a été vérifié par le plus grand nombre des faits."

- Karl Popper 1902-1994

Critère de démarcation entre science et non-science

Est scientifique ce qui est falsifiable, se soumet à falsification empirique = réfutable par les faits, possibilité de l'invalider, de la réfuter ou encore de la tester.

Qu'est-ce que la recherche ? [1] – *Courants de pensées*

- Thomas Kuhn 1922-1996

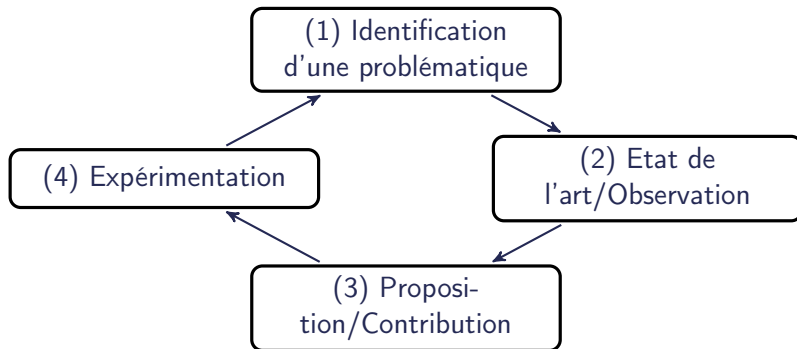
Existence d'un paradigme, d'une matrice disciplinaire

- Généralisations symboliques (= langage commun de la "tribu")
- Croyances communes (ontologie, métaphores et analogies)
- Valeurs communes (valeurs, normes méthodologiques, épistémologiques)

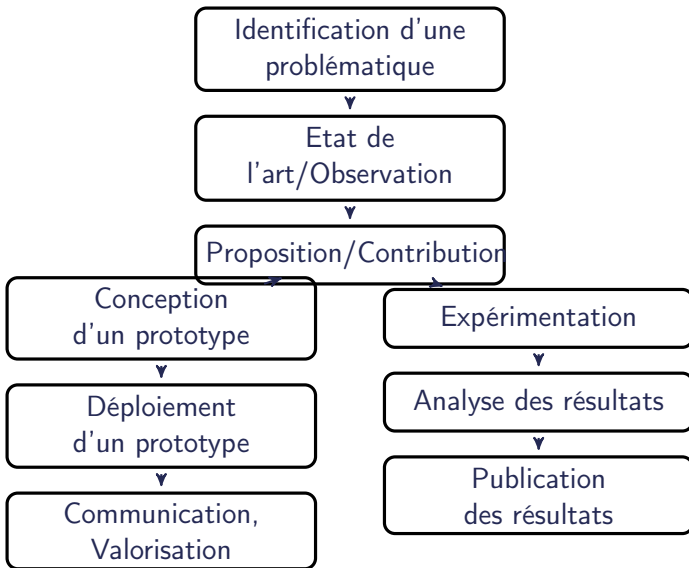
Qu'est-ce que la recherche en informatique ?

- L'informatique
 - Aspects expérimentaux (programmer)
 - Aspects "sciences sociales" (étude du comportement des utilisateurs)
 - Aspects théoriques : démontrez qu'un problème est difficile...
 - Aspect multi-disciplinaire : math-info, info-santé, info-océanographie, ...
- Exemples de recherche en machine learning
 - Etude de la classification dans un très grand nombre de catégories
 - Apprentissage de représentations pour des données générés par les utilisateurs
 - Apprentissage de distance pour la comparaison d'images et de pages Web
 - Modèles séquentiels pour la classification multiclasse, sparse et budgétée
 - Évaluation des systèmes de recommandation à partir d'historiques de données

Processus de recherche 1/2



Processus de recherche 2/2



Exercice :

Énoncé : Un chercheur propose nouveau téléphone portable NEWt, avec un nouveau mode d'interaction : le TouchT. Il pense que ce dispositif est plus performant et plus précis que deux modèles concurrents Oldboy et Oldfriend. Il voudrait connaître les améliorations qu'il doit apporter à l'interface pour la rendre plus facile d'utilisation.

Quelles questions un chercheur doit-il se poser ?

Exercice :

Énoncé : Un chercheur propose nouveau téléphone portable NEWt, avec un nouveau mode d'interaction : le TouchT. Il pense que ce dispositif est plus performant et plus précis que deux modèles concurrents Oldboy et Oldfriend. Il voudrait connaître les améliorations qu'il doit apporter à l'interface pour la rendre plus facile d'utilisation.

Quelles questions un chercheur doit-il se poser ?

- Quels faits allez vous chercher à démontrer ?
- Quelle expérience proposez-vous ? Quelle est la contre expérience ?
- Quelles personnes allez-vous choisir pour faire les expériences ? Pourquoi ?
- Quels types de mesures allez-vous réaliser ? Quelles données allez-vous obtenir ?
- Quel est le matériel expérimental dont vous aurez besoin ?
- Quels types de traitements envisagez vous ?

Projet

Comprendre, proposer, évaluer des modèles

- Réaliser un état de l'art
- Proposer un nouveau modèle
- Proposer un protocole d'évaluation adapté

→ Thématiques autour des modèles de langue → Travail en groupe
(3 personnes OBLIGATOIREMENT !!!)

Pas si simple que ça...

- On reste dans la réflexion et l'organisation
→ Pas d'implémentation !!!!

Thématiques de recherche

<https://arxiv.org/pdf/2303.18223>

- Pré-entraînement des modèles de langue (caractéristiques des données, fonctions de coûts, scalabilité taille architecture avec données/stratégies entraînement,)
- Tokenisation, données en entrées, ...
- Architectures (encoder-decoder, encoder-only, decoder-only...)
- Entraînement et adaptation des modèles de langue
- Ensembling de modèles
- Capacités émergentes des modèles de langue (ICL, COT, variabilités des tâches, quelles connaissances ?, manipulation des outils)
- Importance du prompt dans les LLM (zéro-shot, prompt-tuning,)
- RLHF
- Interprétabilité des modèles de langue
- Frugalité des modèles de langue

Thématiques de recherche

<https://arxiv.org/pdf/2303.18223>

- Gestion des biais et fairness
- Gestion des hallucinations
- Prise en compte des contextes longs
- Stratégie de décodage
- Utilisation de sources de données externes (bases de connaissances)
- Évaluation des modèles de langue
- Évaluation par les modèles de langue
- Gestion des données privées / suppression des données
- Détection de la génération
- Apport de la multimodalité
- Gestion des langue peu dotées/multilinguisme

TODO

Identifier le thème sur lequel vous voulez travailler

Références

[1] http://www.dphu.org/uploads/attachements/books/books_216_0.pdf

[2] http://www.recherche-qualitative.qc.ca/documents/files/revue/edition_reguliere/numero27%282%29/paille27%282%29.pdf

[3] <http://mescal.imag.fr/membres/jean-marc.vincent/index.html/M2R-Methodo/index.html>

[4] http://pagesperso.lina.univ-nantes.fr/~cdlh/Downloads/Methodologie_2014.pdf

[5] <https://explorable.com/fr/methodologie-de-recherche>