

Audition

Mesdames et messieurs les membres du jury,

Je vous remercie de me donner l'opportunité de vous présenter mon projet de recherche, qui a pour ambition de développer une grille d'analyse des effets du numérique par le biais de ses codes sources.

Introduction

L'extrait de code que vous voyez ici est tiré du code source du logiciel *Moodle*, version **1.0.6**. Moodle est un logiciel de gestion de l'éducation (le plus populaire en Europe). En cela, il **formalise et automatise certains aspects de l'enseignement**.

Et je voudrais attirer votre attention sur la ligne numéro 5: `if teacher authority > 0`, qui peut être lue de deux manières.

D'une certaine manière, on peut lire cette ligne comme une **branche conditionnelle**, un des éléments de base de la programmation. Ici, elle décide de l'affichage de variables si un enseignant a une valeur d'autorité qui est supérieure à 0.

D'une autre manière, on peut lire cette ligne comme la **traduction d'un lien entre autorité et enseignant**, rapport qui dénote une hiérarchie inhérente à une certaine version de l'éducation.

Ces deux manières de lire soulèvent plusieurs questions. Quelle est la fonctionnalité de cette autorité comme champ de variable dans le code source de *Moodle*? Qu'est-ce que cette autorité permet de faire, ou empêche de faire, dans un environnement virtuel, et dans le monde réel? En d'autres termes, comment faire sens de la **polysémie** des concepts inscrits dans les codes sources?

Ce sont ces questions sur la signification et l'action du code source qui motivent mon projet de recherche. En effet, **le code informatique est le seul langage qui fait ce qu'il décrit**, puisqu'il peut être exécuté par une machine. Il s'agit pour moi de comprendre comment les logiciels, les langages de programmation, les protocoles ou encore les formats de fichiers vont, dans leur structure, représenter une certaine version du monde, et de comprendre l'influence qu'une telle représentation peut avoir dans le monde.

Projet

Autrement dit, je compte saisir comment les artefacts computationnels encodent différentes versions du monde, et évaluer leurs conséquences.

Je vais commencer par resituer le projet comme étant à la croisée de **multiples enjeux** autour des effets normatifs et prescriptifs du numérique.

Je vais ensuite vous présenter une **méthodologie intégrative**, qui part de l'étude des codes sources.

Je finirai enfin par détailler comment je compte déployer cette méthode à travers des **différents corpus et différents terrains**.

Enjeux

Commençons donc par les enjeux.

Actuellement, les logiciels sont trop souvent traités de “**boîtes noires**”, dont le fonctionnement interne se soustrait à toute lecture critique, et complique toute tentative d’audit. La recherche actuelle a recours à des approches de tests externes qui peuvent rencontrer certaines difficultés et pourraient alors être complétées par d’autres méthodes.

Plus largement, cette conception de “boite noire” nuit à une littéracie rigoureuse du numérique, que l’on peut voir lors de controverses autour de logiciels tels que Parcoursup. Soit cette opacité pose les logiciels comme incompréhensibles et implique que nous y soyons soumis, impuissants; ou, au contraire, cette opacité nous invite à n’étudier que leurs contextes et leurs effets, et non leur structure.

Ce que ce projet propose, c’est **passer de la boîte noire à la grille de lecture**.

D’un point de vue **social et juridique**, l’enjeu est celui d’une double dynamique de numérisation de la société, et d’impératifs de transparence de cette numérisation (au niveau européen comme au niveau national). Pouvoir saisir l’impact des logiciels implique de pouvoir lire et interpréter leurs structures afin de considérer si ces logiciels sont compatibles avec certaines valeurs données (équité, partage du savoir, liberté de choix, etc.).

D’un point de vue **scientifique**, il s’agit de contribuer au débat dans entre les études des médias et les études de sciences et techniques entre une conception **déterministe**, selon laquelle la technologie va forcer des manières de penser et des manières d’agir chez les usagers et usagères, et une conception **constructiviste**, selon laquelle la fonction d’un artefact est toujours changeante selon les contextes et les usages situés.

Enfin, d’un point de vue **pratique**, l’augmentation du nombre de lignes de codes générées par des grands modèles de langages demande que l’on développe des capacités à lire du code, plutôt qu’uniquement à l’écrire.

Méthode

Pour répondre à ces enjeux, je propose une méthodologie intégrative.

La lecture et la compréhension du code est un problème aussi vieux que la programmation elle-même. Les **sciences informatiques** ont pour cela développé des méthodes effectives d'écriture du code et d'analyse automatique. Seulement il s'agit uniquement de **confirmer une adéquation à une spécification donnée**, mais pas d'établir les effets plus larges de son exécution.

Les approches en **sciences humaines**, elles, se concentrent sur la signification culturelle d'extraits de code de taille limitée. Une opération de code, ou un nom de fonction, vont être examinés de manière ponctuelle et à l'aune de contextes culturels très vastes, et non dans des cas d'études précis.

Enfin, les approches en **sciences sociales** vont souvent traiter le code comme un objet opaque, difficile à étudier, et donc se concentrer sur ses modes de production et sur ses modes de réceptions, délaissant sa matérialité spécifique.

Ma méthodologie est intégrative en ce qu'elle réunit ces trois approches.

Fondamentalement, elle s'ancre dans le champ des *software studies*, une interdiscipline croisant littérature et études des médias cherchant à mettre à jour, par des lectures rapprochées de codes sources, par des approches d'archéologie des médias, et par de la recherche-création, quelles sont les spécificités de cette nouvelle forme d'écriture qu'est le code informatique.

Ces *software studies* sont principalement développées dans les mondes anglo-saxons et germanophones, où j'ai longtemps travaillé, mais encore moins présentes en France, même si je compte articuler mes recherches avec des initiatives existantes, comme le séminaire Codes Sources, ou le projet Software Heritage.

En cela, il s'inscrit aussi dans les **humanités numériques**, pas seulement comme application de techniques numériques (classification, quantification, automatisation) pour comprendre un corpus humaniste, mais aussi, et surtout, comme l'application de techniques humanistes (interprétation, herméneutique, contextualisation) pour comprendre un corpus numérique.

Je mobiliserai aussi des études ethnographiques, afin de saisir comment les programmeurs et les programmeuses, ainsi que les usagers et usagères, font sens de leurs pratiques et de leurs ouvrages, notamment en termes d'**imaginaires socio-techniques**, ou de rapport entre **stratégie** et **tactique**.

Cette méthodologie demande donc une capacité bien particulière à naviguer entre les approches des sciences informatiques, sciences humaines et sciences sociales.

Parcours

Ce sont des compétences que j'ai développées à travers mon parcours interdisciplinaire et international, à travers l'Europe, l'Amérique du Nord, l'Asie et la péninsule arabique.

J'ai d'abord acquis une base solide en sciences humaines et sociales lors d'un master en sciences politiques, avant de découvrir les enjeux du numérique à travers un master en design de jeu vidéo. Puis, après un séjour de deux ans dans un département d'arts et médias interactifs, j'ai soutenu une thèse sur les codes sources en littérature comparée.

Plus récemment, mes différents postes de recherches post-doctorales, en philosophie des médias et humanités numériques, m'ont permis de développer une forte autonomie dans les projets que je mène.

En parallèle, j'ai toujours conservé une pratique professionnelle de la programmation, notamment en développant des oeuvres d'art numérique.

Axes de recherche

Afin de répondre à ces enjeux, je compte appliquer cette méthodes à travers différents axes de recherche

d'abord en formalisant des approches d'écriture et de lecture de code pour analyser la structure des logiciels.

puis en comparant les résultats de ces analyses dans les contextes de production et de diffusion de savoir,

avant d'élargir mon objet d'étude à des concepts plus vastes, en me penchant sur les spatio-temporalités du numérique.

Lecture et écriture

À court-terme, il s'agit donc d'établir des méthodes de lecture et d'interprétation.

La lecture des codes sources est une activité extrêmement complexe, notamment parce que **la syntaxe des langages de programmation** ne se superpose que partiellement avec la syntaxe des langages naturels (sans même prendre en compte le fait que ces langages soient essentiellement calqués sur la langue anglaise, et développés dans des environnements culturels très restreintes).

Pour l'instant, ce problème a été adressé par les sciences informatiques en considérant le code comme une **grammaire sans contexte**, et je compte donc la recontextualiser.

Mais surtout, on ne lit pas du code comme on lit un texte. La lecture d'une base de code est une affaire de navigation transversale et spatialisée d'architectures conceptuelles, plutôt que de déroulement séquentiel de phrases en prose.

Ce mode de lecture propre au code pourrait être facilité par un **outil d'analyse statique interactif**, développé par exemple en collaboration avec les équipes du laboratoire FEMTO-ST de l'UTBM. Cela, afin de pouvoir représenter, puis traverser, et enfin interpréter comment un logiciel représente le monde.

Au niveau **sémantique**, j'établirai une **grammaire d'action**, qui serait un ensemble de motifs syntactiques ayant des implications sémantiques, permettant le travail d'interprétation.

Celle-ci serait guidée par des questions établies lors de ma thèse sur les rapports entre esthétique et épistémologie dans les codes sources. Lorsqu'une structure de code source est lue (que ce soit une déclaration de fonction, une implémentation de classe, ou une simple suite d'instructions), il faut se demander ce qui est fait, comment cela est fait, et pourquoi cela est fait.

Cette nécessité est une des conclusions de travaux préliminaires sur le code source de Yandex, équivalent de la combinaison de Google et d'Uber dans le monde russophone. En cela, **ce code source constitue une perspective unique pour comprendre la structure des plateformes de gouvernance algorithmique**. Il est aussi d'une échelle inédite, représentant plus de 44 gigaoctets de texte, tandis que l'état de l'art de l'analyse de code se concentre actuellement sur quelques centaines de kilo-octets au plus.

L'étendue des applications de Yandex (moteur de recherche, systèmes de gestion de la publicité, livraison à domicile ou encore navigation cartographiques) pourrait alors constituer à lui seul l'objet d'un financement de type ANR Jeune Chercheur, notamment afin d'assurer des postes d'ingénieur.es de recherche pour développer ces outils numériques de lecture assistée par ordinateur.

Cet axe de recherche contribuera donc à poser des bases méthodologiques par le biais de l'anatomie d'une plateforme, vue de l'intérieur. Cela prendrait la forme de guides de méthode et de série de publications, ainsi que de logiciels de recherche, tous publiés en libre accès.

Production et diffusion du savoir

Le premier axe me permettra donc d'élaborer une grammaire d'action, une sorte de vision idéale de l'action d'un logiciel sur son champ d'application. **À moyen-terme**, il s'agira ensuite de savoir à quel point cette vision idéale du monde est normative, ou prescriptive.

Puisque la programmation manipule des concepts pour effectuer des actions, le champ d'application qui me semble le plus pertinent pour évaluer cette grammaire d'action serait alors l'enseignement et la recherche.

Dans de récents travaux sur *Moodle*, je montre comment une architecture modulaire et des initialisations de variables par défaut peuvent avoir des conséquences sur l'agentivité de l'individu dans une université numérisée. La modularité, l'assignation de rôle et la visibilité seraient par exemple des éléments de cette grammaire d'action.

Mais une telle analyse structurelle doit être éprouvée par un décalage méthodologique, en effectuant des enquêtes sur deux terrains.

D'abord, au sein d'une université utilisant *Moodle* afin d'étudier comment les éléments de cette grammaire sont perçus, vécus et utilisés par les membres du personnel administratif, du corps enseignant et du corps étudiant.

Ensuite, auprès de la fondation *Moodle*, afin de reconstituer les procédés humains qui ont mené à ces inscriptions techniques spécifiques. Cela permettra alors de souligner les forces et les limites des termes proposés lors de l'analyse structurelle, mais aussi de suggérer le développement de nouveaux termes.

Cette approche pourrait aussi prendre de l'ampleur en étant doublement comparative. Il s'agirait de sélectionner une série de logiciels similaires (ici, environnement numérique de travail), afin de noter comment leurs structures varient face à une fonctionnalité commune, mais aussi de confronter ces conceptualisations avec les manières dont les acteurs et les actrices de l'enseignement s'en emparent dans leur pratique quotidienne. Pour cela, j'ai déjà identifié plusieurs terrains.

À terme, cet axe peut se développer au-delà de l'enseignement, la recherche se voyant aussi reconfigurée par l'irruption d'objets numériques. Le rôle épistémique des logiciels y est indéniable, mais encore sous-défini. Un angle d'approche possible pour saisir **ce que le numérique fait à la science** serait d'étudier les logiciels de modélisation utilisés en SHS.

Par exemple, vous voyez ici le code source du Latin American World Model, un modèle de développement socio-économique du monde, conçu par la fondation Bariloche, et en réponse à un modèle du MIT, World3 (lequel était à la base du rapport Meadows sur les limites à la croissance). Ces deux modèles définissent des structures de données bien différentes: World3 inscrit comme unités de bases la consommation et la démographie, tandis que le Latin American World Model pose comme alternatives le niveau d'éducation ou le prix du mètre carré du logement.

Ici, il faudrait alors se pencher sur la question de **comment certaines valeurs peuvent être réifiées dans les outils de la recherche**.

Spatio-temporalités

Enfin, à **plus long terme**, mon dernier axe de recherche élargira l'horizon de mes objets d'étude afin d'aborder de grandes questions culturelles.

Il part de l'observation que le temps et l'espace sont des **composantes cruciales du vécu humain**, et que le numérique participe d'une réduction des distances et d'une accélération des temporalités.

Pour saisir la spatialité propre à la computation, je commencerai par réfléchir aux différents niveaux d'échelles auxquels opèrent différents langages de programmation, tel que l'échelle locale ou globale, mais aussi aux genre d'échelles qu'ils participent à rendre opérationnelles.

D'autre part, il faudrait s'interroger sur le rapport entre localité et matérialité, entre les langages de haut niveau et de bas niveau, et donc sur la prise en compte du substrat matériel de la computation. Cette question de la matérialité est notamment informée par la **croissance rapide de l'empreinte carbone des technologies numériques**. Il s'agira particulièrement de prendre en compte les problématiques de **relocalisation et de dématérialisation** à l'aune des enjeux de la crise climatique.

Le numérique cherche aussi à dompter le temps en proposant toujours plus de rapidité. Je compterai alors examiner plus en détail comment les artefacts computationnels décrivent et activent une infrastructure de la communication dite de "temps réel". Cela inclurait des approches comparatives des différents protocoles de notifications des systèmes d'exploitation mobile (comme Android et iOS). Cela, afin d'établir comment ces protocoles sont cadrés discursivement, comment ils dépendent d'une certaine infrastructure, mais aussi de mettre à jour de possibles alternatives, notamment à travers des travaux de recherche-création.

Cette question de **l'alternative technologique** pose, en filigrane, la question des alternatives sociales, économiques et politiques qui l'accompagnerait.

Conclusion

En conclusion, puisque ce projet est centré sur un objet plutôt que sur une discipline, il se prête naturellement à des collaborations entre sciences informatiques et SHS, telles qu'on peut les trouver en France dans les universités de technologie. Je compte aussi mobiliser les différents réseaux de recherche dont je fais partie, et en développer de nouveaux, en particulier à l'international, pour élaborer une connaissance commune à partir des expertises de chacun et chacune.

Ce projet de recherche propose donc une approche comparative des artefacts computationnels, conjuguant interprétation de codes sources, et enquêtes de terrain, avec pour ambition de saisir, à terme, les grandes dynamiques résultant de l'insertion et de la diffusion massive du numérique comme technique de traduction opérationnelle du monde.

Je vous remercie de votre attention, et je me réjouis de répondre à vos questions,