

## **Rapport de recherche**

**Titre : CHATBOT – Conception, développement et représentation d'un agent conversationnel-acteur. Paradigme de jeu et dialogue théâtral.**

[\(<https://www.manufacture.ch/fr/4467/Chatbot-jouer-et-dialoguer-avec-un-agent-conversationnel-acteur>\)](https://www.manufacture.ch/fr/4467/Chatbot-jouer-et-dialoguer-avec-un-agent-conversationnel-acteur)

Auteur : Nicolas Zlatoff

Date : Octobre 2021

Institution : La Manufacture – Haute école des arts de la scène / HES-SO

Le projet CHATBOT avait pour enjeu la conception, le développement et la représentation d'une intelligence artificielle (IA) sur le modèle d'un.e acteur.rice qui improviserait à partir d'un texte, comme il.elle a l'habitude de le faire en répétitions. Regroupant 4 acteur.rice.s (Elsa Thebault, Lisa Veyrier, Guillaume Ceppi, Lucas Savioz), 3 informaticiens (Jérémy Wenger, Clément Hongler, Marc Riener), une dramaturge (Claire de Ribeaupierre) et un assistant de recherche (Bartek Sozanski), sous la direction de Nicolas Zlatoff, notre équipe a cherché à construire un espace commun de représentation et de jeu, dans lequel une IA et un groupe d'interprètes seraient capables de dialoguer (produire du texte) ensemble. Comment mettre en tension les altérités d'un.e acteur.rice vivant.e face à une machine dépourvue du désir d'être en scène ?

Au niveau informatique, nous avons travaillé à partir du système open source GPT-2 d'OpenAI (Radford, 2019), qui est une référence dans la génération automatique de texte par apprentissage profond. L'apport fondamental de ce type d'approche réside dans le fait qu'il ne s'agit pas de formaliser (déclarer) explicitement à la machine

certaines procédures d'écriture ou de compréhension du texte, mais à lui en faire apprendre certaines, par l'exemple (Lecun, 2019). Ainsi, dans une première phase d'apprentissage, il s'agit de fournir une sélection de textes au système (réseau de neurones), de manière à ce qu'il apprenne, en ajustant ses paramètres, à générer la suite de différents fragments du corpus d'apprentissage. Une fois l'apprentissage effectué, un.e opérateur.rice humain.e peut soumettre n'importe quel texte au réseau afin que celui-ci en génère une suite probable. Pour ce faire, il ne copie pas des extraits précis de son corpus, mais produit une nouvelle association de lettres, qui « ressembleraient » à ce que le réseau a lu lors de son apprentissage. On constate alors un résultat qui peut paraître surprenant : bien que, à strictement parler, le réseau ne voit que des séries de lettres, sans aucune connaissance explicite de leurs sémantiques, il est capable, le plus fréquemment, de former des mots exacts (qui existent) et de les agencer dans une phrase, syntaxiquement correcte et qui tend à signifier quelque chose. Ces résultats sont possibles eu égard au nombre considérable de paramètres d'apprentissage utilisés (le nombre de neurones du réseau, soit entre 114 millions et 1,5 milliards) et d'itérations d'apprentissage (le nombre de prédictions et de comparaisons avec le texte original : plusieurs centaines de milliers ou millions, soit des semaines ou des mois de calculs), qui conditionnent, de fait, un corpus d'apprentissage d'un volume considérable (40 Go de texte brut, pour le système GPT-2 d'OpenAI).

Afin de concevoir un réseau capable de dialogues en français (et donc, par exemple, de savoir gérer des suites de répliques et notamment de jouer des concepts d'altérité et d'identité comme « je » et « tu »), nous avons procédé à un apprentissage en deux temps. Tout d'abord, nous avons agrégé plusieurs corpus libres de droits et diffusés par la communauté scientifique, contenant principalement l'intégralité des livres en français libres de droits (5,5 Go) et des articles (8 Go) diffusés par Wikipedia, des agrégats automatiques (crawling) issus du web francophone, les archives des Nations Unies et de

l'Union Européenne, pour un volume total d'environ 50 Go<sup>1</sup>. On pourrait dire que le réseau apprend, dans cette étape à « parler français », sans notion de dialogue théâtral, avec une variabilité très large de niveau de langue et de références sémantiques. Dans un deuxième temps, nous avons entraîné le réseau sur un corpus exclusivement dialogué, c'est-à-dire sur des successions de répliques (ces dernières sont explicitement indiquées dans le texte, au moyen de balises que nous avons insérées). Ce corpus comprend environ 600 textes de théâtre, depuis les tragédies grecques et romaines à des pièces du début du XXe siècle, en passant par des textes des XVIe, XVIIe et XVIIIe siècles. En outre, afin que le système soit capable de réagir à du texte issu d'une certaine oralité et/ou contemporanéité, nous avons ajouté des transcriptions d'improvisations menées par des interprètes dans le cadre d'une précédente recherche de l'équipe (Zlatoff, 2020), les cours de Gilles Deleuze à l'Université de Vincennes, ainsi que la transcription de différents échanges que nous avons eus au sein de notre équipe, tout au long du projet. Pour cette deuxième étape, on parle de *fine-tuning*, car le réseau se spécialise, sans oublier totalement l'étape d'apprentissage précédente. On pourrait dire que dans cette deuxième phase le réseau « apprend à créer des dialogues de théâtre », une fois qu'il sait parler français.

Les corpus d'apprentissage sont partagés en ligne, accompagnés de notices explicatives, et de différents scripts qui permettent, par exemple, d'insérer automatiquement les balises indiquant les répliques de dialogues, de lancer les apprentissages, d'interroger le réseau sous différentes modalités, notamment à l'aide d'une interface web multi-utilisateurs.

(Voir <https://zenodo.org/record/4923940> et <https://zenodo.org/record/5145506>)

---

<sup>1</sup> Nous nous sommes basés sur les travaux du groupe de recherche GETALP (Grenoble) qui a entraîné son réseau FlauBERT sur la base de données multilingue OPUS.

Nous mettons également en accès libre les différents réseaux que nous avons entraînés, à savoir trois réseaux respectivement de 114, 355 et 774 millions de paramètres, chacun déclinés en deux versions : l'une issue de la première phase d'apprentissage (générale) et l'autre issue du fine-tuning (orienté dialogue).

Parce que les données manipulées par le réseau sont, in fine, du texte écrit, les premières expérimentations scéniques ont consisté à « placer » les acteur.rice.s sur scène, assis.es à une table, derrière un ordinateur, afin qu'il.elle.s puissent saisir un texte écrit pour le transmettre (l'adresser) au réseau. La réponse de ce dernier était affichée, en temps réel, par vidéo-projection. L'affichage lettre par lettre provoquait une théâtralité saisissante, rendant ainsi visible l'altérité dématérialisée d'une machine pensante : le spectateur « ne peut s'empêcher de projeter une vie émotionnelle ou une sorte de conscience sur ces machines » (Dorsen 2010). Mais comment était-il possible de mettre en scène des acteur.rice.s face à une machine dématérialisée ? Comment mettre en scène les corps et la parole dans ce contexte ? Comment donner une présence théâtrale à une machine qui émet des lignes de codes ? Et surtout, comment le faire dans le temps réel de la scène, celui de l'improvisation, sans script précis, en accueillant les propositions inédites d'un réseau de neurones ?

Plusieurs variantes à ce premier dispositif ont été testées, dans l'optique d'augmenter les interactions entre les interprètes sur scène et le réseau de neurones :

- un.e interprète sur scène prête son corps au réseau pendant que s'affiche au-dessus de lui.elle la projection vidéo du texte qu'il produit. Equipé.e d'une oreillette, l'interprète entend également la synthèse vocale de ce texte (TTS : Text-To-Speech), ce qui lui permet notamment de jouer les didascalies (« d'un geste brusque », « s'asseyant », etc.), et augmente de fait les interactions physiques avec la partenaire humaine.
- un.e interprète sur scène prête son corps et sa voix au réseau : toujours à l'aide de l'oreillette et du TTS, il.elle relaie instantanément à l'oral le texte produit par le

réseau, en plus de jouer les didascalies produites.

Nous avons également mis plusieurs réseaux en dialogue entre eux. Des acteur.rice.s sur scène jouaient instantanément ce dialogue, toujours à l'aide d'un système d'oreillettes et de TTS. Avec ce dispositif, les réseaux sont considérés « uniquement » comme des auteurs, des producteurs de textes que les acteur.rice.s jouent instantanément. Deux variantes ont été testées : l'une en temps réel (les interprètes relaient instantanément à l'oreillette la production des réseaux) et l'autre en différé (les interprètes disposent d'un temps avant le relais, qui leur permet de partager avec le public les modalités de jeu dont il.elle.s vont s'emparer pour relayer la production du réseau).

Le jeu avec la machine met ainsi en évidence les qualités de l'interprète humain.e : sa connaissance du répertoire (personnages, intrigue, style de la pièce convoquée), sa mémoire, son sens de l'improvisation. Il.Elle est aux aguets, sans cesse au travail, analysant ce qui arrive, choisissant la réponse adéquate, proposant, s'adressant au public, cherchant des solutions pour essayer de faire tenir une narration et en garder le fil. La machine, souvent, ignore le contexte : elle saute d'une pièce à l'autre. Elle propose un personnage de tragédie, puis appelle un valet de comédie. Et quand l'un d'eux apparaît, c'est avec tout son contexte. C'est alors à l'acteur.rice de le saisir et de le rendre présent, de le restituer et de lui donner vie, pour que le public puisse s'amuser de ce coup de théâtre, avoir la joie de reconnaître une piste, un signe et de trouver des repères. Le jeu devient celui d'une improvisation permanente, la tâche pour l'interprète étant de maintenir une ligne de cohérence avec ce qu'elle connaît de la scène ou en attend, face aux écarts, sauts, divagations ou incohérences régulièrement introduits par le réseau. Alors que, traditionnellement, il s'agit de jouer avec un partenaire dans le cadre de règles ou protocoles fixés à l'avance, le principe ici consiste à jouer à deviner les règles que la machine utilise pour répondre. L'acteur.rice se trouve déporté.e de sa place traditionnelle vers une position de joueur.euse dans un dispositif, tel celui du jeu vidéo. Tout se passe comme si le dispositif d'interaction avec le réseau redoublait ou

affirmait cette position : plongé.e dans un dispositif de simulation d'une histoire du théâtre, l'acteur.rice peut choisir de redoubler ou au contraire de déposséder les *personnages* (convoqués par le réseau) de certaines de leurs qualités ou traits connus, pour en faire autre chose dans le jeu : des *avatars*, c'est-à-dire comme « des marionnettes, des places vides pour une logique de l'action » (Triclot, 2011).

Nous avons rendu public l'ensemble de ces protocoles scéniques à l'Arsenic, centre d'art scénique contemporain (Lausanne) les 26 et 27 mars 2021. Le format intégrait également, sous une forme didactique, des clés de compréhension du fonctionnement du réseau, pour créer un espace de représentation hybride, entre la scène et la conférence, afin d'exposer à la fois le fonctionnement de la machine mais également celui des interprètes en relation avec elle. Une version plus longue de ce format et qui proposera également au public la possibilité d'interagir avec la machine, sera présenté les 25, 26 et 27 novembre aux Substances (Lyon, France).

La recherche a donné lieu à deux articles dans la presse (numéro 21 de la revue *Hémisphères*<sup>2</sup> ainsi que *le Courrier*<sup>3</sup> du 1<sup>er</sup> avril 2021). Un podcast a été réalisé et est consultable sur le site de la recherche de La Manufacture. Il présente les enjeux de la recherche et des interviews de chacun des membres de l'équipe<sup>4</sup>. En parallèle, un article a été rédigé dans le Journal de la recherche n°2, publié par La Manufacture en janvier 2021<sup>5</sup>. Une soumission sera également faite dans une revue scientifique (relecture en

---

<sup>2</sup><https://www.manufacture.ch/download/docs/e7ryfspm.pdf/DANESI%20Marco,%20«%20Un%20espace%20hybride%20entre%20le%20théâtre%20et%20l'auditoire%20»,%20Revue%20Hémisphères%20n°%2021,%20juin%202021%20.pdf>.

<sup>3</sup><https://www.manufacture.ch/download/docs/rwkf9mrh.pdf/«%20Humain%20Connecté%20»%20%20Le%20Courrier%2001.04.21.pdf>.

<sup>4</sup> [https://soundcloud.com/recherche\\_manufacture/savoirs-sensibles-episode-1-chatbot-de-nicolas-zlatoff](https://soundcloud.com/recherche_manufacture/savoirs-sensibles-episode-1-chatbot-de-nicolas-zlatoff).

<sup>5</sup><https://www.manufacture.ch/download/docs/bfr4ft7v.pdf/Journal%20de%20la%20Recherche%20N°2.pdf>.

double-aveugle). Enfin, les enjeux et les résultats de la recherche ont été présentés par Nicolas Zlatoff à l'INSA (Lyon, France) le 4 février dans le cadre d'un séminaire d'enseignement (filière Télécom, Master 1<sup>ère</sup> année).

Nous avons constaté que ces protocoles scéniques d'interaction entre acteur.rice.s et le réseau de neurones avaient en commun de déplacer les outils traditionnels de jeu des interprètes (notamment le fait de s'adresser au réseau via l'outil du dispositif STT et non plus directement) et esquissaient ainsi la possibilité de construire une collaboration entre un réseau de neurones et les interprètes, une hybridation du jeu d'acteur.rice avec une machine génératrice de texte. Par suite, de nouvelles questions sont apparues : comment tisser des liens, développer une intelligence commune ? Nous avons en effet observé qu'un mode de compétition ne fonctionnait pas avec la machine, puisqu'elle ne partage pas les mêmes compétences que les interprètes. La machine peut-elle être considérée comme un instrument ? Le rôle de l'interprète est-il de bien jouer de cet instrument ? Si la machine n'occupe qu'une position d'*objet*, elle induit une relation peu féconde. Il s'agit d'éviter d'instrumentaliser les instruments, si on veut pouvoir ouvrir de nouveaux champs d'expérience, des relations inattendues. Il s'agirait donc pour l'acteur.rice et l'équipe de chercheur.se.s, d'essayer de saisir au mieux la machine en tant que *sujet*, c'est-à-dire sa « manière d'être au monde », sa façon de communiquer, de proposer, de répondre, de s'écarter. Le comprendre d'un point de vue technique et scientifique, mais également d'un point de vue plus sensible. Pour que l'on puisse jouer ensemble, il faut projeter de l'intentionnalité à la machine, de l'imagination. C'est justement le rôle de la création et du théâtre et cela fait actuellement l'objet d'un nouveau projet de recherche, dont la requête, auprès du FNS, est en cours de rédaction par Nicolas Zlatoff.

DORSEN Annie, Hello Hi There, conception Annie Dorsen, production : Steirischer Herst, Hebbel am Ufer, BIT Teaergarasjen, Black Box theater, PS122, première le 24.09.2010 au Steirischer Herst (Graz).

LE CUN Yann, Quand la machine apprend, la révolution des neurones artificiels et de l'apprentissage profond, Paris, Odile Jacob, 2019.

RADFORD Alec, WU Jeff, CHILD Rewon, LUAN David, AMODEI Dario, SUTSEKEVER Ilya, Language Modelare Unsupervised Multitask Learners, 2019

TRICLOT Mathieu, Philosophie des jeux vidéo, Paris, Zones, 2011.

ZLATOFF, Nicolas, Vers un paradigme d'interprétation-en-jeu d'un texte, au double sens du jeu et des études littéraires (INTERPRETATION) – Rapport d'activité. La Manufacture HETSR, Lausanne, 2020. Disponible sur :

<http://www.manufacture.ch/download/docs/ktbhyxg3.pdf/Rapport%20d'activite%CC%81.pdf> Consulté le 20.01.2021.