

## La vallée dérangement de l'intelligence artificielle

Dans les années 1970, des roboticiens ont dégagé l'idée d'*uncanny valley* : une créature artificielle qui a l'air d'un humain mais pas tout à fait est beaucoup plus dérangement pour nous qu'une qui a l'air plus clairement artificielle ou au contraire une qui est vraiment indiscernable d'un humain. Cette conjecture a largement été confirmée depuis, que ce soit avec de vrais robots ou, encore plus, des robots fictifs. En français, on dit « vallée dérangement » ou « vallée de l'étrange », et le mot vallée fait référence à un creux dans le graphe de la fonction qui relie ressemblance et sympathie.

J'ai l'impression qu'il se produit un phénomène similaire avec l'intelligence artificielle : elle nous dérangement quand elle est trop intelligente mais pas assez pour l'être autant que nous.

Mais quand j'écris ça, j'exploite le fait que le mot dérangement a plusieurs sens : les robots de la vallée de l'étrange nous dérangent comme une idée ou une image dérangement : ils nous mettent mal à l'aise, nous font peur de manière nébuleuse. Alors que le problème dont je veux parler avec l'intelligence artificielle, c'est qu'elle nous dérangement comme un bruit ou du chahut : elle perturbe notre concentration et le flux de notre activité.

Pour commencer, parlons d'ordinateurs, de déterminisme et d'apprentissage.

Un ordinateur est quelque chose de fondamentalement déterministe : si on lui présente plusieurs fois la même entrée, il répondra à chaque fois le même résultat.

Enfin, ça c'est la théorie. Un ordinateur est un dispositif électronique, et à ce titre exposé à toutes les fluctuations inévitables des mécanismes physiques, à fortiori parce qu'il est poussé très près des limites de ses possibilités. C'est pour ça qu'un ordinateur trop overclocké plante fréquemment ; un ordinateur tournant à sa vitesse nominale peut planter aussi pour la même raison, mais beaucoup plus rarement. D'autre part, la communication avec l'extérieur, avec les périphériques, fait intervenir des événements asynchrones dont l'instant précis, mesuré avec toute la précision de l'horloge du processeur, est imprévisible. Donc un ordinateur n'est pas vraiment déterministe.

Mais en pratique, on peut presque faire comme s'il l'était, au moins en première approximation. Les manières évidentes de programmer comptent sur ce déterminisme et ont tendance à gommer le hasard issu des périphériques.

Notons que ce hasard est parfois bienvenu, et pas seulement dans les jeux vidéo. En cryptographie, la clef qui protège une communication doit être choisie au hasard de manière à ce qu'un espion ne puisse pas la deviner.

Pour un utilisateur humain, ce déterminisme est une aubaine. Le cerveau est tout spécialement optimisé pour détecter les déterminismes de ce genre. Et une fois qu'il les a détectés, il fait des prédictions, et utilise ces prédictions pour an-



anticiper. C'est quelque chose de complètement omniprésent : nous anticipons le temps qu'une porte va mettre à s'ouvrir ou se fermer pour passer la main sans nous faire pincer ; nous anticipons que le jet qui va sortir d'une bouteille d'eau va aller vers l'avant pour mettre le verre au bon endroit, etc. C'est quelque chose que nous faisons vraiment très bien.

Et donc nous anticipons les ordinateurs. Ils ont beau être rapides, il y a souvent des éléments lents qu'il faut attendre, à commencer par le disque dur et le réseau. Par exemple, je viens peut-être de cliquer sur un lien pour télécharger un fichier, je sais qu'on va me demander le nom sous lequel je veux l'enregistrer. Je peux anticiper, commencer à taper le nom alors que le téléchargement est toujours en cours. Il y a un petit risque, s'il y a une erreur imprévue, ce que j'ai tapé peut avoir des conséquences irritantes, mais c'est rare, un risque calculé. Et si je décide de le prendre, c'est parce que ça me conduit à une meilleure utilisation de mon temps, et donc une meilleure satisfaction.

Hélas, les informaticiens semblent avoir fait leur mission de ruiner ce déterminisme et de rendre le comportement des ordinateurs difficile à prévoir, aléatoire, voire même erratique. Je vais évoquer trois sources de problèmes.

La première manière dont ils s'y sont pris n'a rien à voir avec l'intelligence artificielle : le marché a conduit à faire des interfaces des ordinateurs des accessoires de mode.

Pour apprendre à utiliser efficacement un ordinateur, il est important que les choses restent globalement inchangées. Si l'option dont on a de temps en temps besoin est toujours accessible sous le même onglet depuis le même menu, on retient progressivement comment la trouver de plus en plus rapidement. Si d'une version à l'autre elle change de place, cet apprentissage ne se produit pas.

Mais d'un point de vue commercial, les quelques nouvelles fonctionnalités ajoutées à un logiciel arrivé à maturité ne sont pas suffisantes pour vendre une nouvelle version. Il faut du spectaculaire. Changer l'aspect graphique de l'interface est un moyen bon marché d'apporter de la nouveauté visible. Si ces changements portaient uniquement sur l'apparence, ce ne serait pas trop grave, même si la perte de repères visuels est déjà un problème. Mais ils s'accompagnent presque toujours de changements cosmétiques de comportement : onglets qui passent du haut au côté ou réciproquement, défilement horizontal ou vertical, etc.

Il faut être honnête, certains changements d'interface améliorent réellement l'ergonomie. Il est difficile de juger l'équilibre optimal entre apporter aux utilisateurs une amélioration et préserver leurs repères. Mais il me semble clair que le rythme effréné que nous impose l'industrie actuellement est bien trop intense.

La deuxième source d'aléa néfaste dans les interfaces vient d'une amplification malencontreuse des phénomènes aléatoires qui se produisent lors d'un fonctionnement normal, couplée à une programmation souvent négligente.



Tout tient à l'impression de réactivité. Malgré leur vitesse, les ordinateurs ne peuvent pas répondre à tout instantanément, donc les interfaces prévoient des astuces pour faire patienter l'utilisateur ou lui permettre de continuer sans attendre.

La solution passe souvent par un comportement asynchrone : l'opération est lancée, son résultat sera visible dans quelques secondes, en attendant l'ordinateur continue à réagir comme d'habitude. On le voit facilement à l'œuvre dans les propositions que les moteurs de recherche affichent souvent au fur et à mesure qu'on tape : si le réseau est rapide, les suggestions s'affichent immédiatement, s'il est lent elles mettent du temps à arriver ; mais dans tous les cas on peut continuer à taper.

Ça a l'air très bien, mais imaginez que vous décidez d'arrêter de taper pour choisir une des suggestions juste au moment où de nouvelles suggestions arrivent : vous croyez cliquer sur une chose mais c'est une autre qui se glisse sous la souris.

Ce n'est pas le seul exemple. Si on lance en même temps deux applications également gourmandes, on ne peut pas facilement prévoir celle qui s'affichera en premier, surtout si elles résident sur le même disque mécanique. Mais l'ordre d'affichage aura aussi pour conséquence le placement des fenêtres : la première arrivée aura la place de premier choix à l'écran, la seconde prendra les restes, ou bien au contraire passera devant alors qu'on s'apprêtait à utiliser la première.

Les inconvénients de ces phénomènes asynchrones sont souvent amplifiés par la négligence des développeurs en ce qui concerne le focus. Le focus, c'est le curseur clignotant qui indique quelle ligne du formulaire va recevoir les touches qu'on tape, mais ce n'est pas tout, c'est aussi une ombre grisée ou un liséré en pointillé sur le bouton qui réagira si on appuie sur espace ou le lien qui s'ouvrira si on appuie sur entrée. Il suit les clics de la souris, mais on peut aussi le déplacer au clavier, avec la touche tabulation en général pour le faire passer à l'élément suivant, parfois aussi avec les touches fléchées.

Un comportement déterministe du focus est indispensable à une utilisation efficace et confortable, car la manipulation au clavier est plus précise que la manipulation à la souris. Pourtant, il arrive régulièrement que le focus se perde dans les limbes à l'occasion d'une opération, souvent une opération faisant intervenir quelque chose d'asynchrone. Par exemple, je connais plusieurs sites web où le fait de cliquer sur un bouton qui charge de l'information supplémentaire rend ensuite impossible de défiler dans la page avec les flèches du clavier.

J'en arrive à la troisième raison que je voulais évoquer pour les comportements erratiques des ordinateurs : l'intelligence artificielle, ou plus exactement son utilisation irréfléchie.

Avant toute chose, je tiens à insister sur un fait : l'intelligence artificielle n'est pas quelque chose de nouveau. Quand on plaçait dans les fusées V-2 un accéléromètre connecté à deux circuits intégrateurs pour qu'elles évaluent leur altitude,



c'était déjà de l'intelligence artificielle. Très rudimentaire, certes, mais de l'intelligence artificielle néanmoins. Quand une boîte de dialogue fait une régression linéaire sur la vitesse d'un traitement pour nous donner une estimation du temps qui reste, c'est encore de l'intelligence artificielle. C'en serait plus si elle utilisait une vraie méthode des moindres carrés couplée à une moyenne glissante plutôt qu'une bête extrapolation sur les deux dernières mesures, mais ça compte. Ce qu'on fait de nos jours sous cette appellation utilise des données beaucoup plus volumineuses et des traitements beaucoup plus complexes, mais le principe reste le même, il n'y a pas eu de progrès qualitatif important depuis.

L'application de l'intelligence artificielle aux interfaces informatiques consiste à essayer de deviner ce que l'utilisateur veut, ce qu'il s'apprête à faire, pour lui présenter de la manière la plus accessible possible. Tel quel, ça a l'air bien. Et ponctuellement, au cas par cas, ça l'est. Mais l'utilisation d'un ordinateur n'est pas ponctuelle.

On peut utiliser un ordinateur efficacement parce qu'on peut prévoir, anticiper, ses réactions. L'intelligence artificielle rend son comportement beaucoup plus complexe, donc beaucoup plus dur à anticiper. Si l'intelligence artificielle était assez bonne, elle proposerait à coup sûr ce qu'on souhaite, il n'y aurait qu'à valider. Mais ça n'existe pas, ce genre de précision relève de la télépathie ou de la divination, pas de l'intelligence, artificielle ou pas.

Le problème réside ici dans deux acteurs qui cherchent à anticiper mutuellement ce que va faire l'autre. C'est une situation qui revient fréquemment en cas d'adversité, lors de duels fictifs par exemple qui peuvent donner des scènes très réussies, drôles ou spectaculaires. Mais dans la réalité, ça a plus tendance à ressembler à deux personnes qui essaient en vain de se céder le passage de manière symétrique.

Je vais détailler un exemple concret. Un des logiciels que j'utilise fréquemment propose de deviner la fin de ce que j'ai tapé. Ici, « deviner » veut dire supposer que j'ai tapé le début d'un nom de fichier et proposer les fichiers qui sont effectivement présents et dont le nom correspond. Comme c'est un système que j'utilise beaucoup et dont je n'ai pas changé la configuration significativement depuis des années, je sais sans réfléchir exactement quoi taper pour désigner un fichier donné, et encore plus quand je dois taper plusieurs noms de fichiers similaires.

Ce mécanisme peut être rendu plus intelligent en ajoutant des exceptions : une opérations sur un répertoire ne proposera que les répertoires, une commande pour lire des vidéos ne proposera pas les fichiers de code source, etc. Utilisé à bon escient, c'est très pratique. Mais parfois ça a des ratés. Il arrive fréquemment qu'on ait plusieurs fichiers dont le nom ne diffère que par un suffixe ; dans ce cas, taper leurs noms commence par les mêmes touches. Mais un jour, quelqu'un a remarqué que quand il s'agit de supprimer des fichiers, écrire deux fois le même est inutile, et a utilisé cette observation pour rendre la saisie plus intelligente : si un



fichier est déjà dans la liste des fichiers à supprimer, on ne le propose pas. Hélas, cette optimisation rompt la régularité du comportement, et peut amener à de la confusion quand les touches tapées n'amènent pas au résultat attendu.

Je cite cet exemple parce qu'il m'est familier : il m'est effectivement arrivé de m'embrouiller dans la saisie de noms de fichiers à supprimer précisément à cause de cette optimisation. Je pourrais citer d'autres exemples du même genre. Le cœur du problème, c'est que j'avais anticipé que l'ordinateur complèterait, mais je n'avais pas anticipé qu'il omettrait certains fichiers ; et de l'autre côté, l'ordinateur avait anticipé que je voulais taper des noms de fichiers, mais pas que j'avais anticipé comment il complèterait. De plus, le bénéfice apporté par cette optimisation est minuscule, et ne justifie pas l'effort qu'il m'a fallu pour retenir cette exception.

J'ai un autre exemple, peut-être plus familier du grand public. Quelqu'un à l'aise avec un clavier tape vite, mais fait parfois des fautes de frappe, qu'il « sent » et corrige immédiatement. Quelqu'un moins à l'aise tape moins vite, mais la disposition du clavier constante lui permet de retrouver les touches rapidement. Confrontons ces personnes à une borne tactile pour acheter des tickets de train : la borne est intelligente, à chaque lettre, elle n'active sur le clavier que les touches qui correspondent à la suite du nom d'une station qui existe. Pour une personne pas du tout à l'aise, c'est une aide. Pour une personne moyennement à l'aise, les changements d'apparence vont parasiter les repères visuels et ralentir la recherche de la bonne lettre. Pour une personne très à l'aise, le comportement inhabituel, et même imprévisible (car il faudrait connaître la liste de toutes les stations et être capable de la filtrer mentalement très rapidement), peut rendre les choses très inconfortables : en cas de faute de frappe, la touche n'aura peut-être pas été prise en compte, mais les automatismes conduisent à l'effacer quand même.

On pourrait multiplier les exemples. Pour les systèmes grand public, je pense qu'il y a une tendance générale qui ressort : l'intelligence artificielle a été déployée non pas pour compléter la compétence humaine mais pour la suppléer. Ça correspond à un nivellement par le bas, voire pire si des interférences se produisent comme dans les exemples que j'ai donnés.

Pour utiliser un ordinateur efficacement, il faut se construire un modèle mental fidèle de son état, l'incorporer à la proprioception étendue que j'ai évoquée dans « Mon appartement e(s)t moi ». Cependant, si on utilise un ordinateur, c'est souvent pour lui demander des tâches pour lesquelles on n'a pas les capacités mentales, et donc qu'on ne peut pas modéliser. Ça n'a de chance de marcher que si les résultats fournis par l'ordinateur sont gardés un minimum isolés de ce qui fait son comportement vis-à-vis de l'utilisateur. Ce n'est pas le cas actuellement, et à ma connaissance peu de gens pensent au problème en ces termes.

Il y a une boutade qui dit : « je déteste les ordinateurs, ils font toujours ce que je leur dis, jamais ce que je veux » ; la vallée dérangement de l'intelligence artificielle,



c'est quand les ordinateurs ne sont toujours pas assez intelligents pour faire ce qu'on veut, mais se croient trop intelligents pour faire ce qu'on dit.

*Publié le 4 mars 2019*

