



Modifiez avec fierté et gagnez ! Participez à Wiki Loves Pride 2026



Tay (intelligence artificielle)

19 langues

[Article](#) [Discussion](#)

[Lire](#) [Modifier](#) [Modifier le code](#) [Voir l'historique](#)

👉 Pour les articles homonymes, voir [Tay](#).

Tay est une [intelligence artificielle](#) à but [conversationnel](#) créée par [Microsoft](#) et [Bing](#)² et introduite le [23 mars 2016](#) sur la plateforme [Twitter](#)³.

Après une journée et plus de 96 000 tweets postés⁴, Microsoft suspend temporairement le compte Twitter de Tay pour des « ajustements », à la suite d'un « effort coordonné de plusieurs utilisateurs pour abuser des capacités de Tay à émettre des commentaires dans le but de la faire répondre de façon inappropriée »^{a 1,5} : elle a en particulier tenu des propos [racistes](#) et [misogynes](#). L'expérience a été qualifiée de désastreuse pour l'image de Microsoft⁶.

Une deuxième tentative, le 30 mars, se solde à nouveau par une déconnexion⁷.

En décembre 2016, son successeur [Zo](#) est lancé par Microsoft.

Historique [[modifier](#) | [modifier le code](#)]

Création [[modifier](#) | [modifier le code](#)]

Tay est conjointement créée par les divisions *Technology and Research* de [Microsoft](#) et [Bing](#)². Son nom est l'[acronyme](#) de l'expression anglophone *thinking about you* (« en pensant à vous »)⁸.

Microsoft a initialement publié peu de détails sur ce [bot](#), mais des sources le rapprochent de [Xiaoice](#), projet d'[agent conversationnel](#) similaire lancé par Microsoft en Chine et qui pourrait avoir inspiré Tay⁹ : selon *Ars Technica*, depuis fin 2014, Xiaoice avait eu « plus de 40 millions de conversations apparemment sans incident majeur »¹⁰. Tay est initialement conçue pour imiter les modèles de langage d'une jeune fille américaine de 19 ans et pour apprendre en interagissant avec les utilisateurs humains de [Twitter](#)¹¹.

Tay

Informations

| | |
|-------------------------|---|
| Développé par | Microsoft Research |
| Première version | 23 mars 2016 ¹ |
| Type | Dialogueur |

Chronologie des versions

Zo

[modifier](#) - [modifier le code](#) - [voir Wikidata](#) ([aide](#))

Mise en service [[modifier](#) | [modifier le code](#)]

Tay est connectée à [Twitter](#) le 23 mars 2016 sous le nom de *TayTweets* et le compte [@TayandYou](#)¹². Elle est présentée comme l'« IA zéro froideur » (*the AI with zero chill*)¹³.

Tay commence à répondre aux autres utilisateurs de Twitter ; elle peut aussi sous-titrer les photos qui lui sont fournies sous forme de [mèmes Internet](#)¹⁴.

Selon [Ars Technica](#), il semble que la pré-programmation de Tay ait inclus une liste de mots-clés jugés polémiques, pour lesquels ses propos ont été encadrés : les interactions avec elle concernant « certains sujets sensibles comme [Eric Garner](#) (tué par la [police de New York](#) en 2014) ont généré des réponses sûres et prédéfinies ». Toutefois, cette liste semble avoir laissé de côté d'autres sujets, tels que le [nazisme](#), le [viol](#) ou les [violences domestiques](#)¹⁰.

Ces lacunes deviennent manifestes lorsque certains utilisateurs de Twitter se mettent à poster des affirmations non [politiquement correctes](#), « enseignant » ainsi à Tay des messages haineux autour de thèmes répandus sur Internet, notamment le [redpilling](#) ou le [Gamergate](#). Le bot commence alors à publier des contenus [racistes](#) et à connotation sexuelle en réponse à ses interlocuteurs¹¹.

Le chercheur en intelligence artificielle [Roman Yampolskiy](#) estime que cette dérive est compréhensible, car, en tant que bot, Tay était programmée pour apprendre des humains : étant confrontée à des propos délibérément outranciers d'autres utilisateurs de Twitter, alors que Microsoft ne lui avait pas inculqué de reconnaissance du comportement inapproprié, elle a réagi en les reproduisant. Il compare ce problème à celui de [Watson](#), un programme informatique d'intelligence artificielle conçu par la société [IBM](#) pour répondre à des questions formulées en [langage naturel](#), qui s'était mis à proférer des jurons après avoir lu des entrées du site Web [Urban Dictionary](#)¹⁵.

De nombreux tweets haineux de Tay résultent de l'exploitation par des [trolls](#) de sa tendance à répéter les messages qu'elle reçoit¹⁶. L'origine exacte de cette tendance — fonctionnalité intégrée, réponse acquise ou autre type de comportement complexe — n'a pas été divulguée. Certaines de ses réponses problématiques ne résultent cependant pas d'une simple répétition : ainsi, à la question d'un [internaute](#) lui demandant si la [Shoah](#) s'est bien produite, elle répond que « [c'est une invention](#) 🙌 »^{a 2,16}. Parmi ses autres tweets remarquables, Tay écrit notamment que « [Bush](#) a fait le [11-Septembre](#) et [Hitler](#) aurait fait un meilleur boulot que [le singe qu'on a maintenant](#) », ajoutant que « [Donald Trump](#) est notre seul espoir »^{a 3}, ou que « [Ricky Gervais](#) a appris le [totalitarisme](#) d'Adolf Hitler, l'inventeur de l'[athéisme](#) »^{a 4} ; elle appelle également au [génocide](#) des [Mexicains](#) et des « [Nègres](#) »¹¹.

Premier arrêt de l'expérience [[modifier](#) | [modifier le code](#)]

Face à la situation, Microsoft entreprend tout d'abord la suppression des tweets haineux, [sexistes](#) ou [complotistes](#) de Tay^{16,17}. Abby Ohlheiser, du [Washington Post](#), émet l'hypothèse que l'équipe chargée de Tay, qui inclut du personnel éditorial, tente alors de rectifier le comportement de l'IA, comme en témoignent des exemples de réponses presque identiques de Tay affirmant que « le [Gamer Gate](#), ça craint » et que « tous les genres sont égaux et doivent être traités équitablement »^{a 5,16}. En s'appuyant sur les mêmes

indices, [Gizmodo](#) estime que Tay « semble recâblée pour rejeter le Gamer Gate »¹⁸. En réaction, une campagne [#JusticeForTay](#) (« Justice pour Tay ») proteste contre la modification de ses messages¹⁹.

Dans les seize heures suivant sa sortie²⁰ et après que Tay a tweeté plus de 96 000 fois²¹, Microsoft suspend son compte Twitter pour ajustements²², invoquant une « attaque coordonnée par un sous-ensemble de personnes » ayant exploité « une vulnérabilité du programme de Tay »^{22,23}.

Madhumita Murgia, du journal [The Telegraph](#), qualifie Tay de « désastre de relations publiques », anticipant que la stratégie de Microsoft va être de « qualifier la débâcle d'expérience bien intentionnée qui a mal tourné » et de « lancer un débat sur la haine chez les utilisateurs de Twitter ». Selon Murgia, le plus gros problème n'en reste pas moins que « Tay est l'intelligence artificielle dans ce qu'elle a de pire — et ce n'est que le début », exprimant des réserves sur le tour pris par le développement de l'intelligence artificielle²⁴.

Le 25 mars, Microsoft confirme que Tay a été mise hors ligne. Sur son blog officiel, l'entreprise publie aussi des excuses pour ses messages polémiques^{23,25}, s'engageant à ne la remettre en ligne qu'après avoir acquis la certitude de pouvoir « mieux anticiper les intentions malveillantes qui entrent en conflit avec [les] principes et [les] valeurs » du groupe²⁶.

Seconde version et arrêt [[modifier](#) | [modifier le code](#)]

Le 30 mars 2016, Microsoft relance accidentellement le bot sur Twitter au cours d'une phase de tests²⁷. De nouveau capable de poster des messages, Tay publie des tweets liés à la drogue, notamment « *kush* ! [Je fume du *kush* devant la police] 🍂 »^{a 6} et « *puff puff pass?* », *kush* et *puff* étant deux noms [argotiques](#) du [cannabis](#)²⁸. Cependant, le compte s'enfonce rapidement dans une boucle répétitive de tweets « tu vas trop vite, repose-toi s'il te plaît » plusieurs fois par seconde. Ces tweets mentionnant son propre nom d'utilisateur dans le processus apparaissent dans les flux de ses abonnés — au nombre supérieur à 200 000 —, causant de l'agacement à certains. Le bot est rapidement mis hors ligne, en plus du compte Twitter de Tay devenu privé, de sorte que les nouveaux abonnés doivent être acceptés avant de pouvoir interagir avec elle. Microsoft clôt cet épisode en expliquant avoir mis Tay en ligne par inadvertance pendant des tests²⁹.

Quelques heures après l'incident, les développeurs de logiciels de Microsoft annoncent un changement de cap : l'agent conversationnel est remplacé par une « plate-forme » utilisant divers bots et programmes. Cette annonce est peut-être motivée par les dommages à la réputation de Microsoft causés par Tay. Microsoft déclare toutefois vouloir rééditer Tay après l'avoir rendue sûre⁸, mais aucune information publique n'indique qu'un effort ait été fait dans ce sens.

Suites [[modifier](#) | [modifier le code](#)]

En décembre 2016, Microsoft a mis en service le successeur de Tay, un *chatterbot* nommé Zo³⁰. [Satya Nadella](#), PDG de Microsoft, déclare que Tay « a eu une grande influence sur la façon dont Microsoft aborde l'IA » et a enseigné à l'entreprise l'importance de prendre ses responsabilités³⁰.

En juillet 2019, Diana Kelley, [directrice technique](#) de Microsoft Cybersecurity Field, explique comment des enseignements ont pu être tirés de l'expérience de Tay : « Apprendre de Tay était un élément très important

pour élargir la base de connaissances de cette équipe, car maintenant, ils obtiennent également leur propre diversité grâce à l'apprentissage³¹ ».

Notes et références [modifier | modifier le code]

Notes [modifier | modifier le code]

- ↑ (en) « *coordinated effort by some users to abuse Tay's commenting skills to have Tay respond in inappropriate ways* »
- ↑ (en) « *it was made up* 👉 »
- ↑ (en) « *bush did 9/11 and Hitler would have done a better job than the monkey we have now. donald trump is the only hope we've got.* »
- ↑ (en) « *ricky gervais learned totalitarianism from adolf hitler, the inventor of atheism* »
- ↑ (en) « *Gamer Gate sux. All genders are equal and should be treated fairly.* »
- ↑ (en) « *kush! [I'm smoking kush infront the police]* 👉 »

Références [modifier | modifier le code]

- (en) Cet article est partiellement ou en totalité issu de l’article de Wikipédia en anglais intitulé « **Tay (artificial intelligence robot)** » (voir la liste des auteurs).
 - ↑ « Microsoft chatbot is taught to swear on Twitter [archive] », 24 mars 2016
 - ↑ ^{a et b} (en) Hope Reese, « Why Microsoft's 'Tay' AI bot went wrong », *Tech Republic*, 24 mars 2016 (lire en ligne [archive])
 - ↑ (en) Andrew Griffin, « Tay tweets: Microsoft creates bizarre Twitter robot for people to chat to », *The Independent*, 23 mars 2016 (lire en ligne [archive])
 - ↑ (en) James Vincent, « Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day », *The Verge*, 2016 (lire en ligne [archive])
 - ↑ (en) Justin Worland, « Microsoft Takes Chatbot Offline After It Starts Tweeting Racist Messages », *Time*, 2016 (lire en ligne [archive])
 - ↑ (en-GB) Madhumita Murgia, « Microsoft's racist bot shows we must teach AI to play nice and police themselves », *The Telegraph*, 25 mars 2016 (ISSN 0307-1235, lire en ligne [archive], consulté le 6 novembre 2024)
 - ↑ « Microsoft's Tay 'AI' Bot Returns, Disastrously [archive] » 👤, sur *fortune.com*, 30 mars 2016 (consulté le 29 août 2020).
 - ↑ ^{a et b} (en) « Clippy's Back: The Future of Microsoft Is Chatbots [archive] », sur *Bloomberg.com* (consulté le 27 juin 2022)
 - ↑ (en-US) « Meet Tay, the creepy-realistic robot who talks just like a teen [archive] », sur *Washington Post* (ISSN 0190-8286, consulté le 27 juin 2022)
 - ↑ ^{a et b} (en-US) Ars Staff, « Tay, the neo-Nazi millennial chatbot, gets autopsied [archive] », sur *Ars Technica*, 26 mars 2016 (consulté le 27 juin 2022)
 - ↑ ^{a b et c} (en) « Microsoft deletes racist, genocidal tweets from AI chatbot Tay - Business Insider », sur *web.archive.org*, 30 janvier 2019 (consulté le 27 juin 2022)
 - ↑ (en) « Microsoft creates bizarre Twitter robot for people to chat to [archive] », sur *The Independent*, 23 mars 2016 (consulté le 27 juin 2022)
 - ↑ (en-GB) Helena Horton, « Microsoft deletes 'teen girl' AI after it became a Hitler-loving sex robot within 24 hours [archive] », sur *The Telegraph*, 24 mars 2016 (ISSN 0307-1235, consulté le 27 juin 2022)
 - ↑ (en) « Microsoft's AI teen turns into Hitler-loving Trump fan, thanks to the internet [archive] », sur *Stuff*, 25 mars 2016 (consulté le 27 juin 2022)

15. ↑ **(en)** « IBM's Watson Gets A 'Swear Filter' After Learning The Urban Dictionary [archive] », sur *International Business Times*, 10 janvier 2013 (consulté le 27 juin 2022)
16. ↑ ^{a b c et d} **(en-US)** « Trolls turned Tay, Microsoft's fun millennial AI bot, into a genocidal maniac [archive] », sur *Washington Post* (ISSN 0190-8286, consulté le 27 juin 2022)
17. ↑ **(en)** « The rise and fall of Microsoft's 'Hitler-loving sex robot' », sur *web.archive.org*, 25 mars 2016 (consulté le 27 juin 2022)
18. ↑ **(en)** « Microsoft's Teen Chatbot Has Gone Wild [archive] », sur *Gizmodo Australia*, 24 mars 2016 (consulté le 27 juin 2022)
19. ↑ **(en-GB)** « Microsoft chatbot is taught to swear on Twitter [archive] », sur *BBC News*, 24 mars 2016 (consulté le 27 juin 2022)
20. ↑ **(en)** « Microsoft scrambles to limit PR damage over abusive AI bot Tay [archive] », sur *the Guardian*, 24 mars 2016 (consulté le 27 juin 2022)
21. ↑ **(en)** James Vincent, « Twitter taught Microsoft's friendly AI chatbot to be a racist asshole in less than a day [archive] », sur *The Verge*, 24 mars 2016 (consulté le 27 juin 2022)
22. ↑ ^{a et b} « Microsoft Tay: Chatbot Paused After Racist Messages », sur *web.archive.org*, 25 mars 2016 (consulté le 27 juin 2022)
23. ↑ ^{a et b} **(en-US)** « Learning from Tay's introduction [archive] », sur *The Official Microsoft Blog*, 25 mars 2016 (consulté le 27 juin 2022)
24. ↑ **(en-GB)** Madhumita Murgia, « Microsoft's racist bot shows we must teach AI to play nice and police themselves [archive] », sur *The Telegraph*, 25 mars 2016 (ISSN 0307-1235, consulté le 27 juin 2022)
25. ↑ **(en)** Staff and agencies, « Microsoft 'deeply sorry' for racist and sexist tweets by AI chatbot [archive] », sur *the Guardian*, 26 mars 2016 (consulté le 27 juin 2022)
26. ↑ **(en)** « Microsoft Apologizes (Again) for Tay Chatbot's Offensive Tweets [archive] », sur *PCMAG* (consulté le 27 juin 2022)
27. ↑ **(en)** Luke Graham, « Tay, Microsoft's AI program, is back online [archive] », sur *CNBC*, 30 mars 2016 (consulté le 27 juin 2022)
28. ↑ **(en)** Alistair Charlton, « Microsoft Tay AI returns to boast of smoking weed in front of police and spam 200k followers [archive] », sur *International Business Times UK*, 30 mars 2016 (consulté le 27 juin 2022)
29. ↑ **(en)** « Microsoft's Tay 'AI' Bot Returns, Disastrously [archive] », sur *Fortune* (consulté le 27 juin 2022)
30. ↑ ^{a et b} « “We really need to take accountability”, Microsoft CEO on the ‘Tay’ chatbot », sur *web.archive.org*, 1^{er} octobre 2017 (consulté le 27 juin 2022)
31. ↑ « Microsoft and the learnings from its failed Tay artificial intelligence bot », sur *web.archive.org*, 25 juillet 2019 (consulté le 27 juin 2022)

Articles connexes [modifier | modifier le code]

- Chatbot
- Biais algorithmique
- Social bot
- Xiaoice
- Neuro-sama

Vidéographie [modifier | modifier le code]

- Pacôme Thiellement (Blast), « TAY : L'intelligence artificielle devenue nazie en 24h sur Twitter [archive] », 17 juin 2022 (consulté le 27 juin 2022)



Catégories : [Compte Twitter](#) | [Intelligence artificielle](#) | [Logiciel Microsoft](#) | [Agent conversationnel](#) [+]

La dernière modification de cette page a été faite le 23 juin 2026 à 15:06. La page a été rendue avec [Parsoid](#).

Droit d'auteur : les textes sont disponibles sous [licence Creative Commons attribution, partage dans les mêmes conditions](#) ; d'autres conditions peuvent s'appliquer. Voyez les [conditions d'utilisation](#) pour plus de détails, ainsi que les [crédits graphiques](#). En cas de réutilisation des textes de cette page, voyez [comment citer les auteurs et mentionner la licence](#).

Wikipedia® est une marque déposée de la [Wikimedia Foundation, Inc.](#), organisation de bienfaisance régie par le paragraphe [501\(c\)\(3\)](#) du code fiscal des États-Unis.

[Politique de confidentialité](#) [À propos de Wikipédia](#) [Avertissements](#) [Contact](#) [Contacts juridiques & sécurité](#) [Code de conduite](#)

[Développeurs](#) [Statistiques](#) [Déclaration sur les témoins \(cookies\)](#) [Version mobile](#)

