

L'INTELLIGENCE ARTIFICIELLE, SON HISTOIRE, SES PROMESSES, SES SUCCÈS, SES ÉCHECS (ET LES PIÈGES À ÉVITER)

Albrecht Zimmermann

Ramène ta science

Bibliothèque d'Hérouville Saint-Clair, 11/01/25



GREYC

Laboratoire de recherche en sciences du numérique



Normandie Université



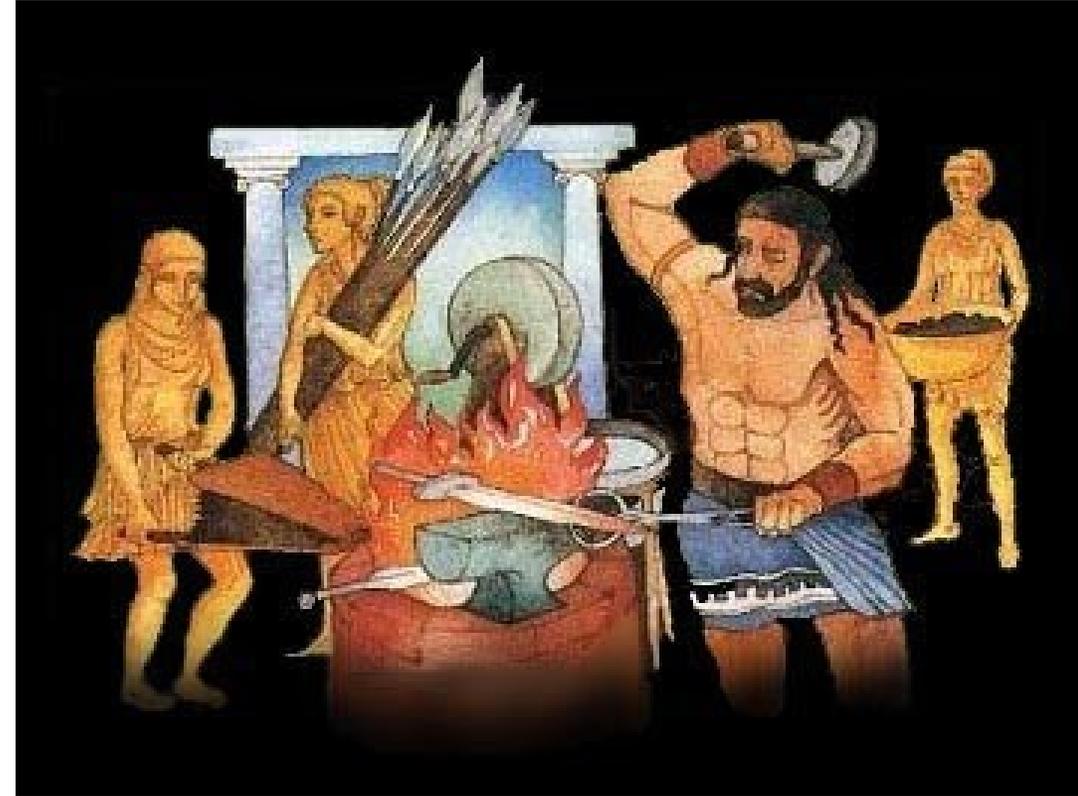
**ENSI
CAEN**
ÉCOLE PUBLIQUE D'INGÉNIEURS
CENTRE DE RECHERCHE



Le rêve, le mythe des êtres artificiels

1. Héphaïstos : Kourai Khrysea (16^{ème} – 12^{ème} siècle av. J.-C.)
2. Juda Loew ben Bezalel : Le Golem (17^{ème} siècle)
 - Devient amok
3. Shelley : Dr. Frankenstein (1818)
 - Prend sa vengeance sur son créateur
4. Karel Čapek : Robot (1920)
 - Révolte → extinction d'humanité → remplacer par les robots
 - Robota (Tchèque) : travail (des serfs)
 - Работа (Russe) : travail

On va revenir sur cette idée



Le Golem

1. Du matériel : l'argile
2. Un programme écrit : le nom de Dieu
3. Travail pour son créateur

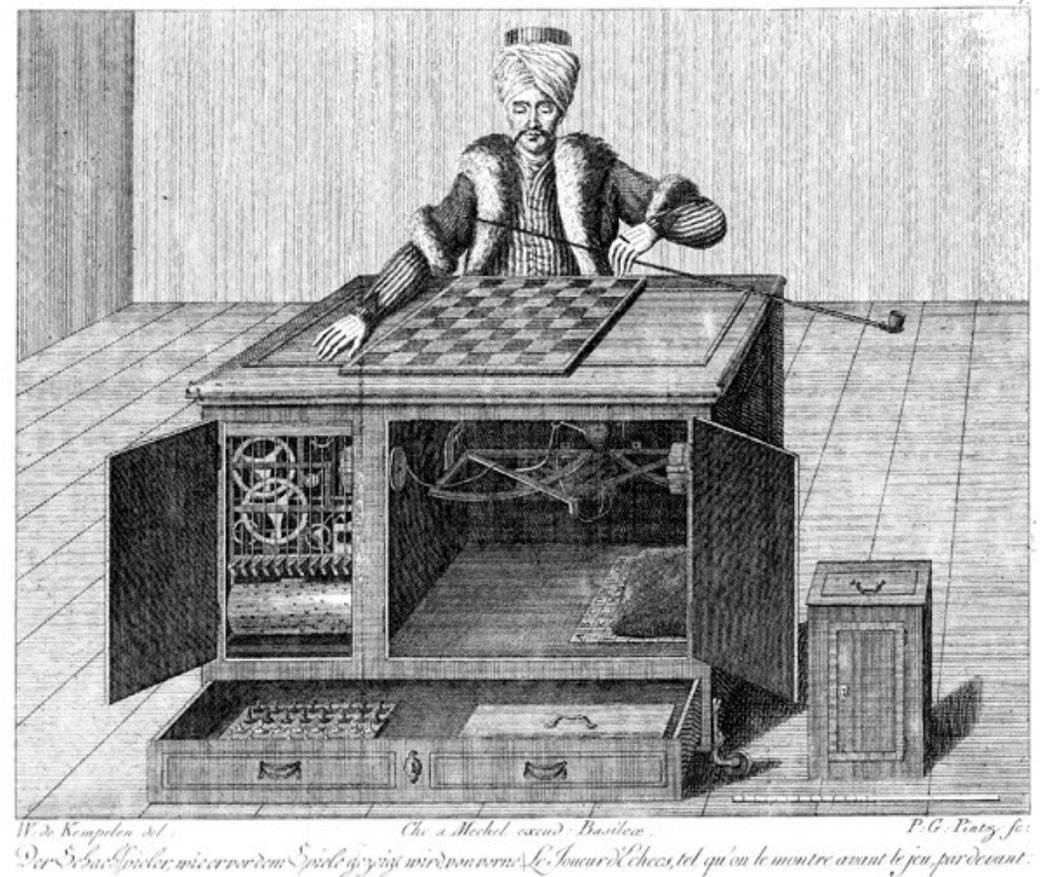


Turc mécanique

- Automaton fort en échecs
- En réalité un joueur humain caché dedans

On va aussi revenir sur ce fait

- Repris par Amazon comme nom pour son service de crowdsourcing



Le début de l'IA - Conférence de Dartmouth (1956)

- Claude Shannon – théorie d'information
- John McCarthy – propose le nom « intelligence artificielle », développe Lisp et Algol, le récupérateur de mémoire
- Nathaniel Rochester – architecte principale du IBM 701
- Marvin Minsky – co-fondateur du labo d'IA au MIT

- Alan Turing « Computing Machinery and Intelligence » (1950)
 - Le jeu d'imitation/le Test de Turing
- Warren McCulloch, Walter Pitts – perceptron (1943)

Optimisme débridé

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. »

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. » 

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. »



Aujourd'hui,
on est proche

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. »
- 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. »



Aujourd'hui,
on est proche

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. » 
- 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. » 

Aujourd'hui,
on est proche

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. »
- 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. »



Aujourd'hui,
on est proche



Deep Blue vaincre
Kasparov en 1997

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. » ❌ **Aujourd'hui, on est proche**
- 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. » ❌ **Deep Blue vaincre Kasparov en 1997**
- 1958 : « Dans dix ans, un ordinateur numérique découvrira et prouvera un nouveau théorème mathématique important. »

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. » ❌ **Aujourd'hui, on est proche**
- 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. » ❌ **Deep Blue vaincre Kasparov en 1997**
- 1958 : « Dans dix ans, un ordinateur numérique découvrira et prouvera un nouveau théorème mathématique important. » ❌

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. » ❌ **Aujourd'hui, on est proche**
- 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. » ❌ **Deep Blue vaincre Kasparov en 1997**
- 1958 : « Dans dix ans, un ordinateur numérique découvrira et prouvera un nouveau théorème mathématique important. » ❌ **Découverte d'un nouveau théorème à l'aide de l'IA en 2021**

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. » ❌ **Aujourd'hui, on est proche**
- 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. » ❌ **Deep Blue vaincre Kasparov en 1997**
- 1958 : « Dans dix ans, un ordinateur numérique découvrira et prouvera un nouveau théorème mathématique important. » ❌ **Découverte d'un nouveau théorème à l'aide de l'IA en 2021**
- 1965 : « Les machines seront capables, d'ici vingt ans, de faire n'importe quel travail qu'un homme peut faire. »

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. » ❌ **Aujourd'hui, on est proche**
- 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. » ❌ **Deep Blue vaincre Kasparov en 1997**
- 1958 : « Dans dix ans, un ordinateur numérique découvrira et prouvera un nouveau théorème mathématique important. » ❌ **Découverte d'un nouveau théorème à l'aide de l'IA en 2021**
- 1965 : « Les machines seront capables, d'ici vingt ans, de faire n'importe quel travail qu'un homme peut faire. » ❌

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. » ❌ **Aujourd'hui, on est proche**
- 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. » ❌ **Deep Blue vaincre Kasparov en 1997**
- 1958 : « Dans dix ans, un ordinateur numérique découvrira et prouvera un nouveau théorème mathématique important. » ❌ **Découverte d'un nouveau théorème à l'aide de l'IA en 2021**
- 1965 : « Les machines seront capables, d'ici vingt ans, de faire n'importe quel travail qu'un homme peut faire. » ❌
- 1967 : « D'ici une génération... le problème de la création d'une « intelligence artificielle » sera en grande partie résolu. »

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. » ❌ **Aujourd'hui, on est proche**
- 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. » ❌ **Deep Blue vaincre Kasparov en 1997**
- 1958 : « Dans dix ans, un ordinateur numérique découvrira et prouvera un nouveau théorème mathématique important. » ❌ **Découverte d'un nouveau théorème à l'aide de l'IA en 2021**
- 1965 : « Les machines seront capables, d'ici vingt ans, de faire n'importe quel travail qu'un homme peut faire. » ❌
- 1967 : « D'ici une génération... le problème de la création d'une « intelligence artificielle » sera en grande partie résolu. » ❌ **Une génération = 30 ans → 1997**

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. » ❌ **Aujourd'hui, on est proche**
- 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. » ❌ **Deep Blue vaincre Kasparov en 1997**
- 1958 : « Dans dix ans, un ordinateur numérique découvrira et prouvera un nouveau théorème mathématique important. » ❌ **Découverte d'un nouveau théorème à l'aide de l'IA en 2021**
- 1965 : « Les machines seront capables, d'ici vingt ans, de faire n'importe quel travail qu'un homme peut faire. » ❌
- 1967 : « D'ici une génération... le problème de la création d'une « intelligence artificielle » sera en grande partie résolu. » ❌
- 1970 : « Dans trois à huit ans, nous disposerons d'une machine dotée de l'intelligence générale d'un être humain moyen. » ❌ **Une génération = 30 ans → 1997**

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. »  **Aujourd'hui, on est proche**
- 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. »  **Deep Blue vaincre Kasparov en 1997**
- 1958 : « Dans dix ans, un ordinateur numérique découvrira et prouvera un nouveau théorème mathématique important. »  **Découverte d'un nouveau théorème à l'aide de l'IA en 2021**
- 1965 : « Les machines seront capables, d'ici vingt ans, de faire n'importe quel travail qu'un homme peut faire. » 
- 1967 : « D'ici une génération... le problème de la création d'une « intelligence artificielle » sera en grande partie résolu. » 
- 1970 : « Dans trois à huit ans, nous disposerons d'une machine dotée de l'intelligence générale d'un être humain moyen. »  **Une génération = 30 ans → 1997**

AGI

Optimisme débridé

- Dans les années 1950 : « Le problème de la traduction automatique sera résolu d'ici cinq ans. »  **Aujourd'hui, on est proche**
 - 1958 : « Dans dix ans, un ordinateur numérique sera le champion du monde d'échecs. »  **Deep Blue vaincre Kasparov en 1997**
 - 1958 : « Dans dix ans, un ordinateur numérique découvrira et prouvera un nouveau théorème mathématique important. »  **Découverte d'un nouveau théorème à l'aide de l'IA en 2021**
 - 1965 : « Les machines seront capables, d'ici vingt ans, de faire n'importe quel travail qu'un homme peut faire. » 
 - 1967 : « D'ici une génération... le problème de la création d'une « intelligence artificielle » sera en grande partie résolu. » 
 - 1970 : « Dans trois à huit ans, nous disposerons d'une machine dotée de l'intelligence générale d'un être humain moyen. »  **Une génération = 30 ans → 1997**
- AGI** 

Chaque été (d'IA) est suivi d'un hiver (d'IA)



Chaque été (d'IA) est suivi d'un hiver (d'IA)

Échec de la
traduction
automatique

e



1956

2024

Chaque été (d'IA) est suivi d'un hiver (d'IA)

Échec de la
traduction
automatique

e



1956

2024

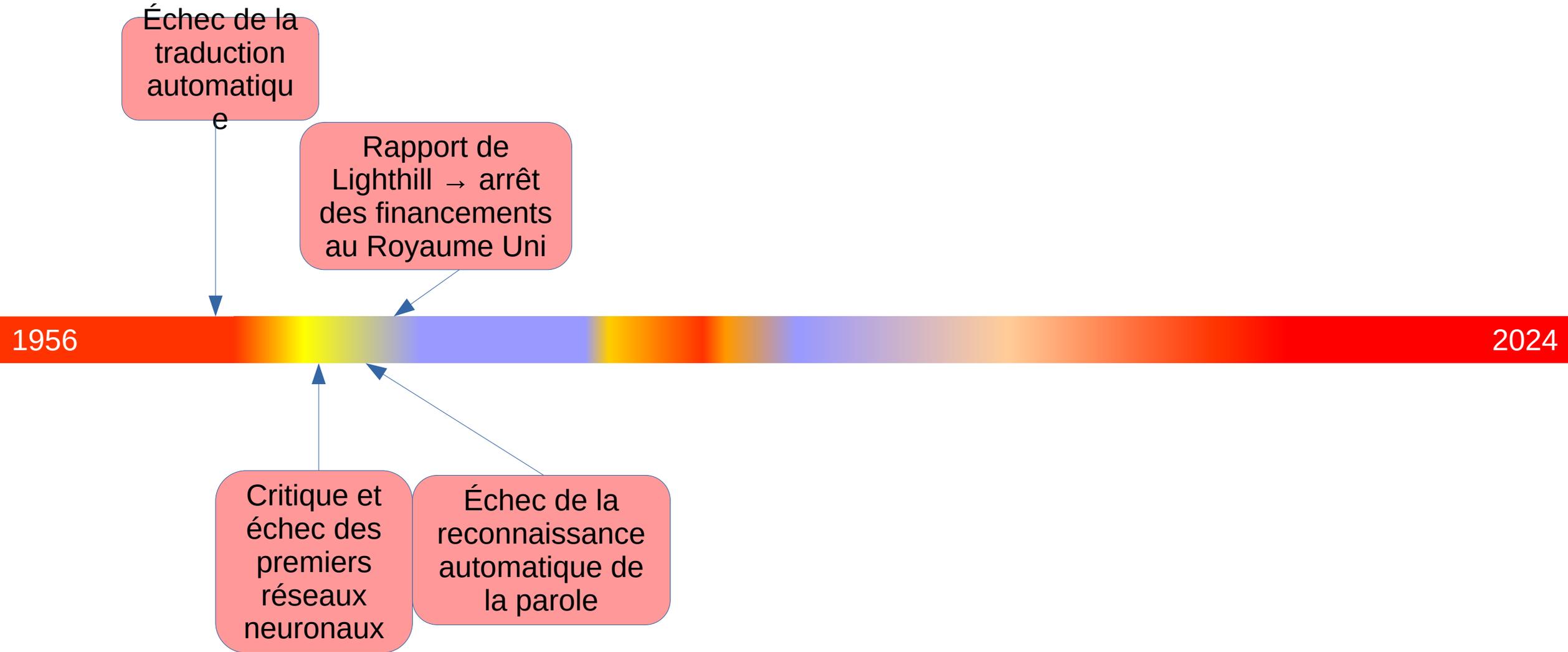
Critique et
échec des
premiers
réseaux
neuronaux



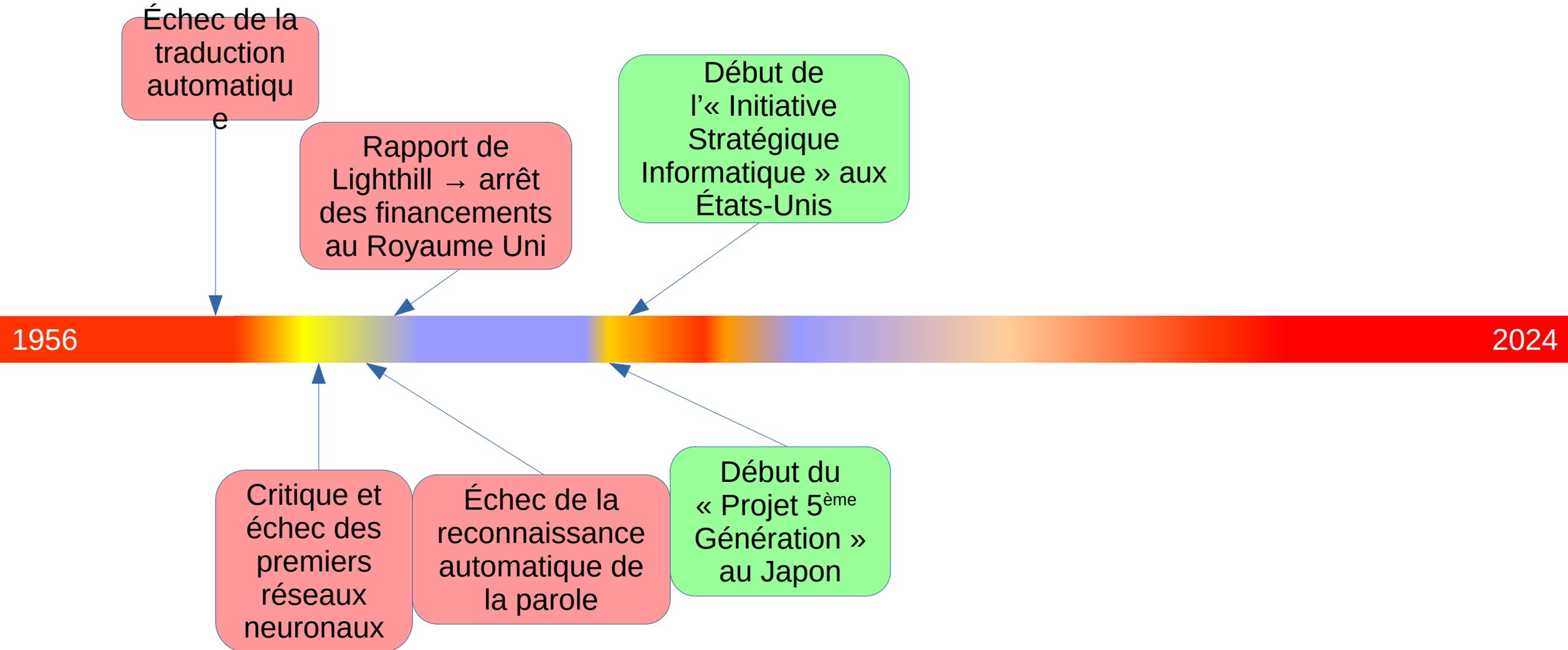
Chaque été (d'IA) est suivi d'un hiver (d'IA)



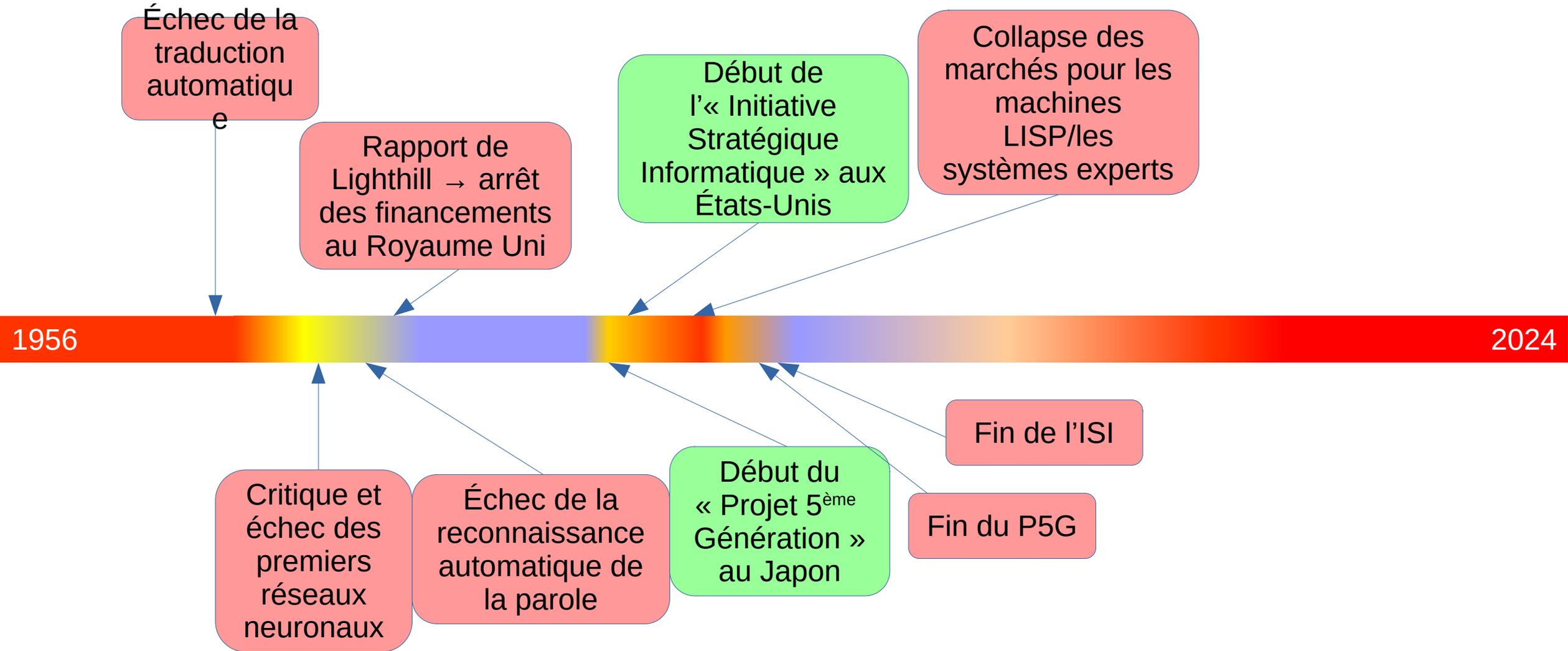
Chaque été (d'IA) est suivi d'un hiver (d'IA)



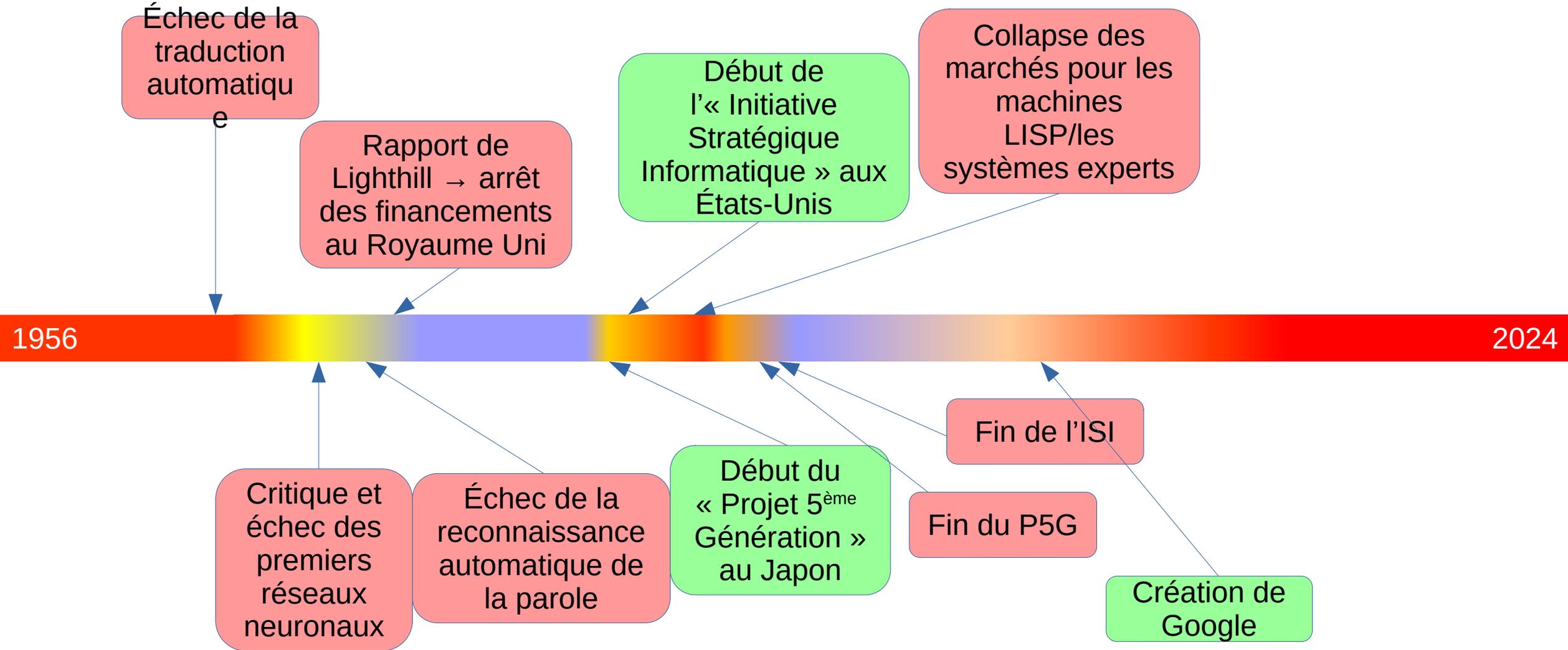
Chaque été (d'IA) est suivi d'un hiver (d'IA)



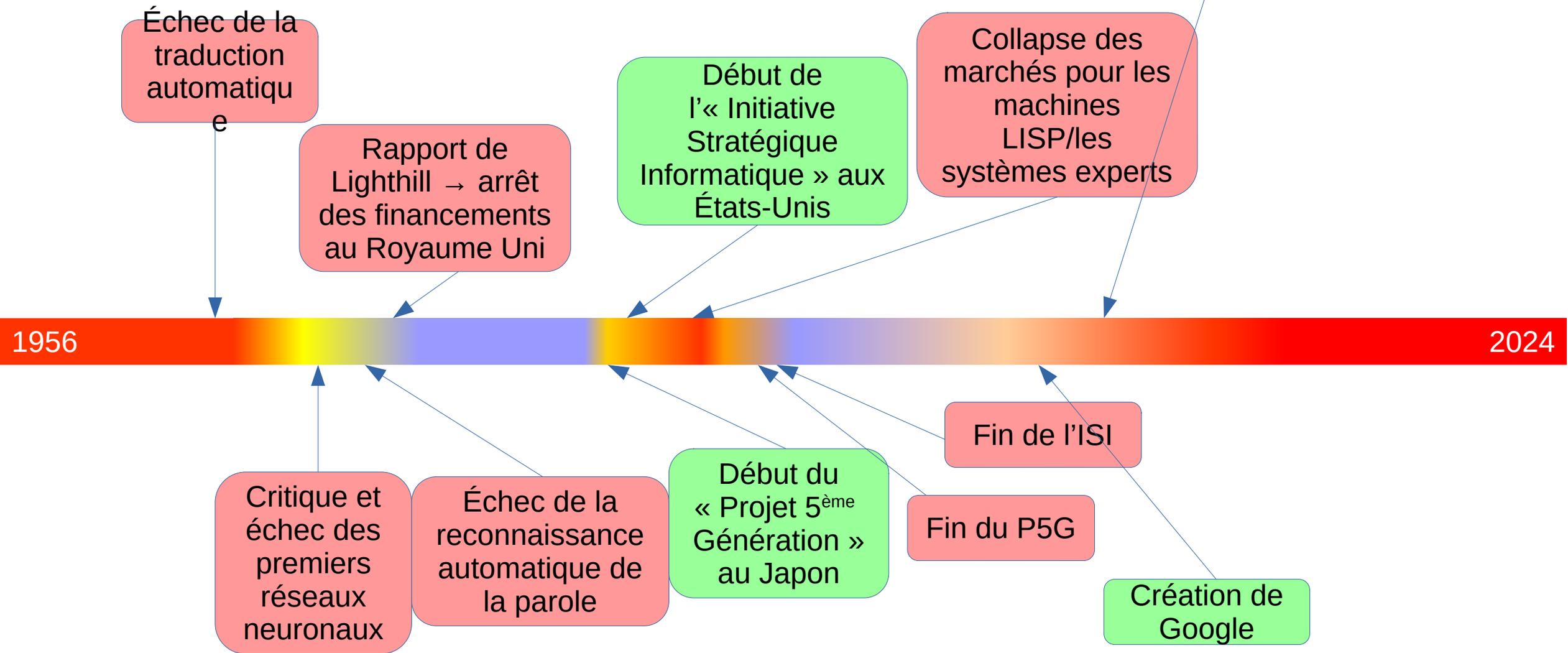
Chaque été (d'IA) est suivi d'un hiver (d'IA)



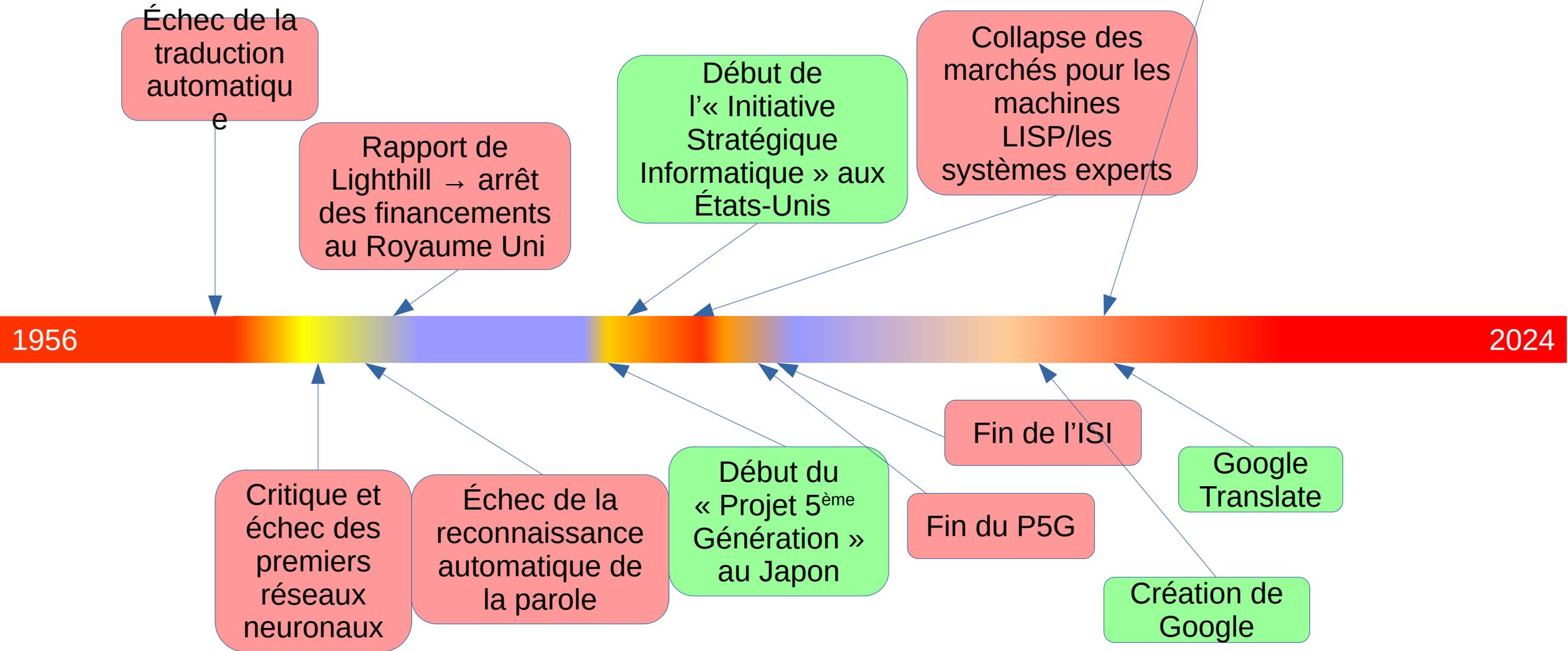
Chaque été (d'IA) est suivi d'un hiver (d'IA)



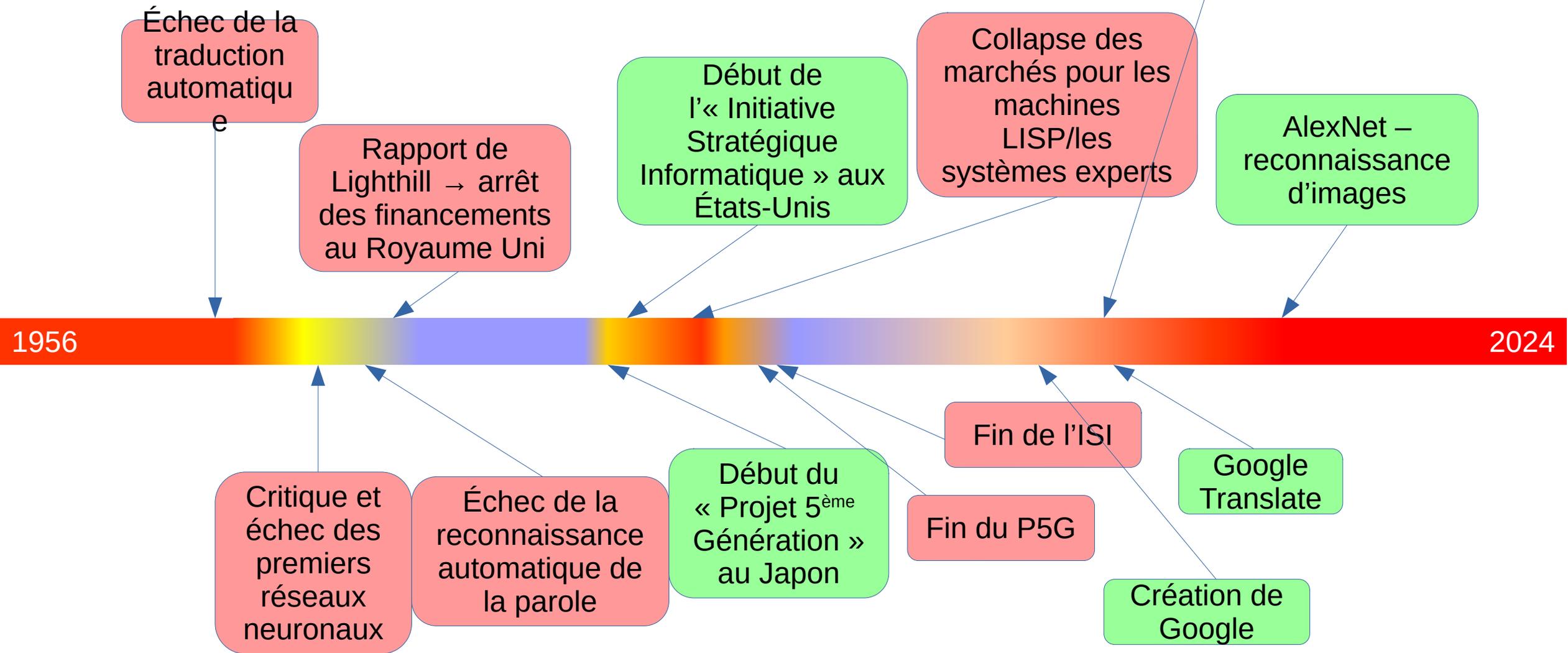
Chaque été (d'IA) est suivi d'un hiver (d'IA)



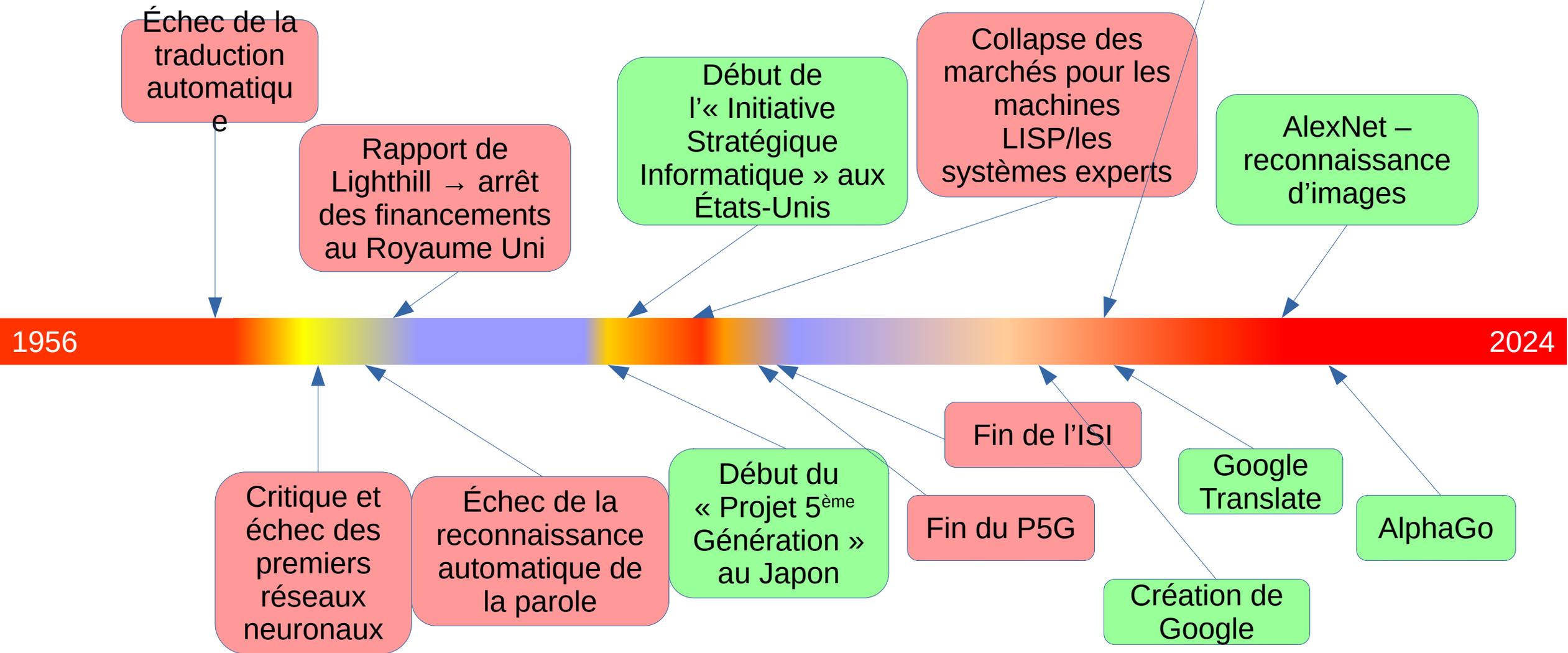
Chaque été (d'IA) est suivi d'un hiver (d'IA)



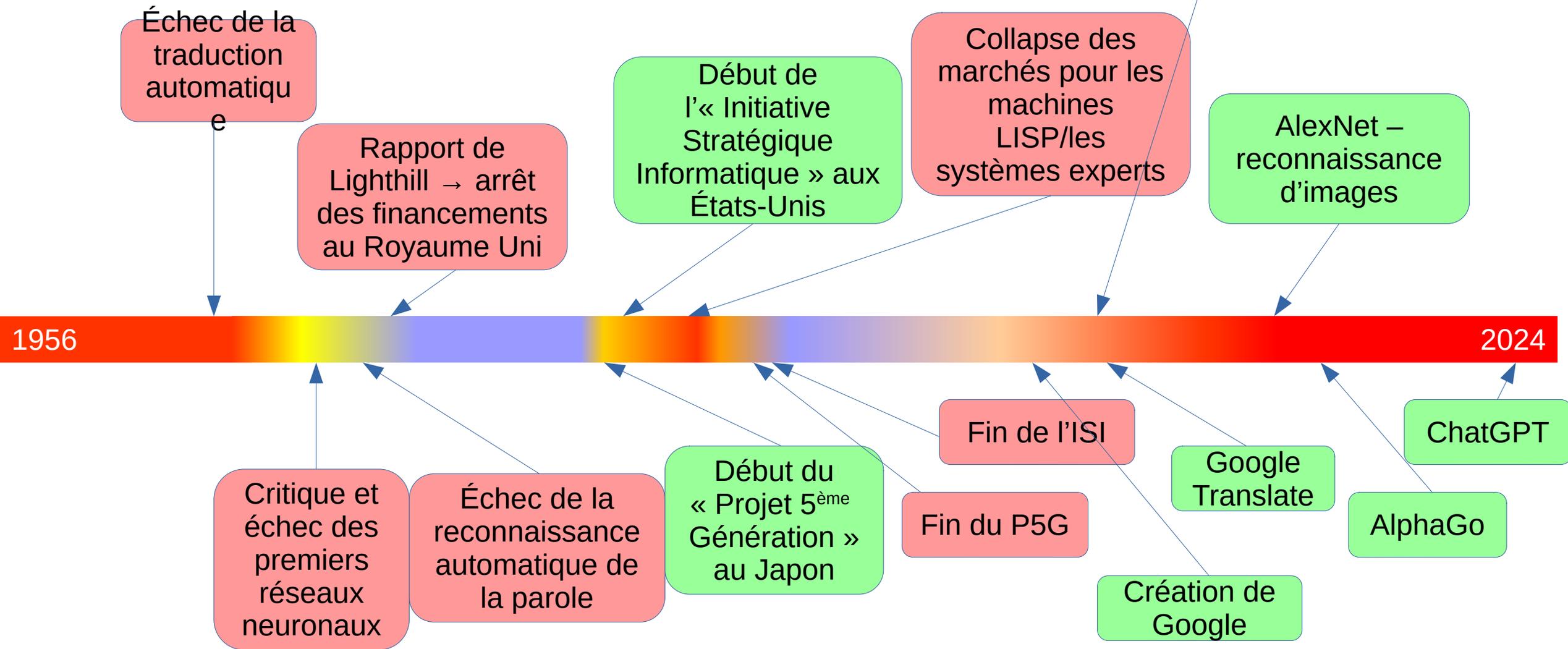
Chaque été (d'IA) est suivi d'un hiver (d'IA)



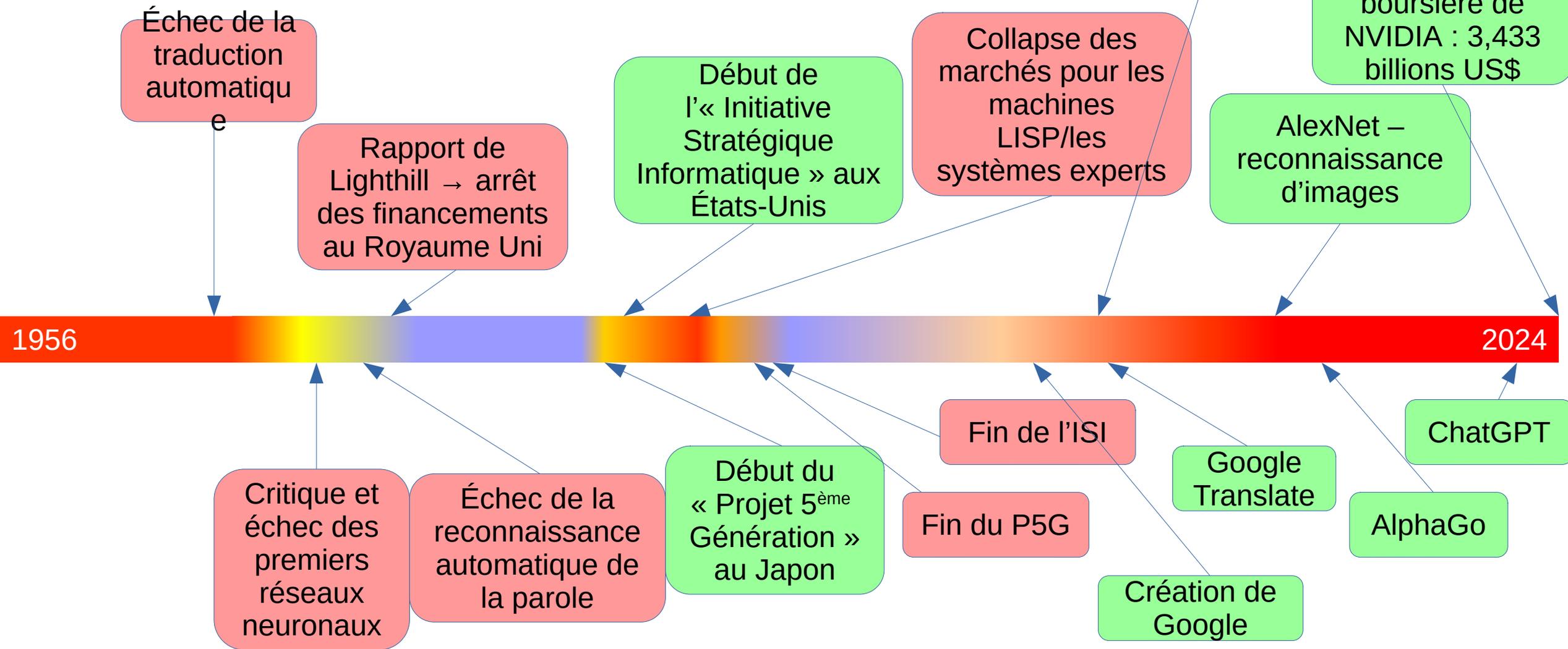
Chaque été (d'IA) est suivi d'un hiver (d'IA)



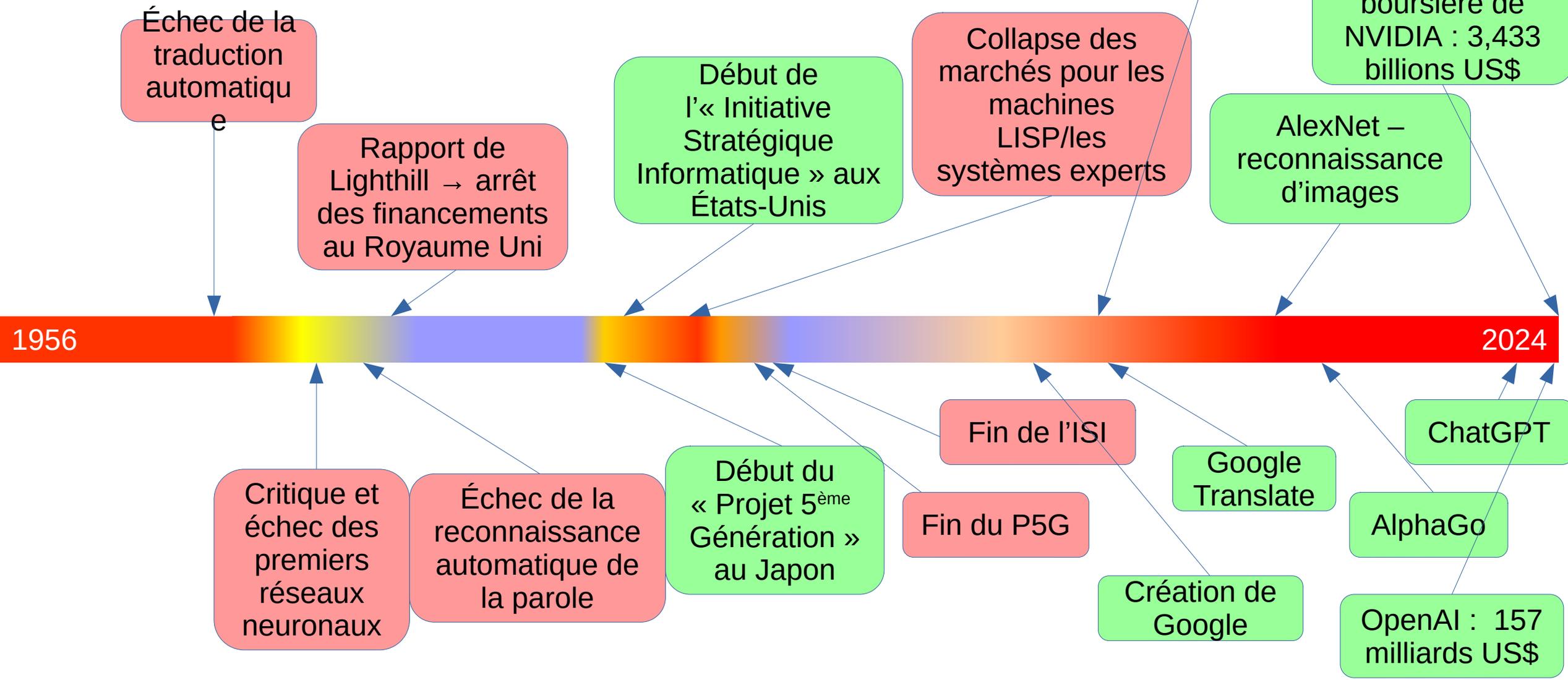
Chaque été (d'IA) est suivi d'un hiver (d'IA)



Chaque été (d'IA) est suivi d'un hiver (d'IA)



Chaque été (d'IA) est suivi d'un hiver (d'IA)



Des raisons pour les échecs

- La paradoxe de Moravec : réussir au raisonnement de haut niveau (démontrer un théorème, jouer aux échecs) \nRightarrow réussir aux aptitudes sensorimotrices
- Application des techniques à des tâches pour lesquelles elles ne sont pas adaptées :
 - Difficile de raisonner avec des réseaux neuronaux
 - Difficile de reconnaître des images avec des règles

Des raisons pour les échecs (2)

- Explosion combinatoire/taille immense de l'espace de recherche
 - Espace de toutes les phrases infinie, par exemple
- Besoin d'immenses masses de faits/données/informations
 - Non pas forcément liés au domaine de la tâche
- Beaucoup des cas exceptionnels
- Puissance computationnelle/stockage insuffisant.e

Cependant, toujours aussi optimiste

- **Elon Musk, Tesla :**
 - 2015 : « Les véhicules Tesla se conduiront tout seuls dans deux ans »
 - 2022 : « Je serais choqué si Tesla ne parvenait pas à atteindre cette année une conduite entièrement autonome plus sûre que les conducteurs humains. »
 - 4, 2024 : « Je suppose que nous aurons une IA plus intelligente que n'importe quel humain probablement vers la fin de l'année prochaine. »
- **Sam Altman, OpenAI :**
 - 9, 2024 : « Il est possible que nous ayons une superintelligence (ASI) dans quelques milliers de jours (!) ; cela peut prendre plus de temps, mais je suis convaincu que nous y parviendrons »
 - 11, 2024 : « OpenAI est en bonne voie pour atteindre l'intelligence artificielle générale (AGI) dans les 5 prochaines années avec le matériel actuel. »
- **Dario Amodei, Anthropic**
 - 2024 : « L'AGI sera atteint d'ici 2026 ou 2027 »

Cependant, toujours aussi optimiste

- **Elon Musk, Tesla :**
 - 2015 : « Les véhicules Tesla se conduiront tout seuls dans deux ans » 
 - 2022 : « Je serais choqué si Tesla ne parvenait pas à atteindre cette année une conduite entièrement autonome plus sûre que les conducteurs humains. »
 - 4, 2024 : « Je suppose que nous aurons une IA plus intelligente que n'importe quel humain probablement vers la fin de l'année prochaine. »
- **Sam Altman, OpenAI :**
 - 9, 2024 : « Il est possible que nous ayons une superintelligence (ASI) dans quelques milliers de jours (!) ; cela peut prendre plus de temps, mais je suis convaincu que nous y parviendrons »
 - 11, 2024 : « OpenAI est en bonne voie pour atteindre l'intelligence artificielle générale (AGI) dans les 5 prochaines années avec le matériel actuel. »
- **Dario Amodei, Anthropic**
 - 2024 : « L'AGI sera atteint d'ici 2026 ou 2027 »

Cependant, toujours aussi optimiste

- **Elon Musk, Tesla :**
 - 2015 : « Les véhicules Tesla se conduiront tout seuls dans deux ans » 
 - 2022 : « Je serais choqué si Tesla ne parvenait pas à atteindre cette année une conduite entièrement autonome plus sûre que les conducteurs humains. » 
 - 4, 2024 : « Je suppose que nous aurons une IA plus intelligente que n'importe quel humain probablement vers la fin de l'année prochaine. »
- **Sam Altman, OpenAI :**
 - 9, 2024 : « Il est possible que nous ayons une superintelligence (ASI) dans quelques milliers de jours (!) ; cela peut prendre plus de temps, mais je suis convaincu que nous y parviendrons »
 - 11, 2024 : « OpenAI est en bonne voie pour atteindre l'intelligence artificielle générale (AGI) dans les 5 prochaines années avec le matériel actuel. »
- **Dario Amodei, Anthropic**
 - 2024 : « L'AGI sera atteint d'ici 2026 ou 2027 »

Pourquoi cet optimisme ?

- Difficile de définir « intelligence » !
- Quelques définitions :
 - « L'intelligence mesure la capacité d'un agent à atteindre ses objectifs dans un large éventail d'environnements. »
 - « [Dans des circonstances nouvelles pour lesquelles l'instinct, l'apprentissage passé ou l'habitude ne dispose d'aucune solution] Aptitude à appréhender et organiser les données de la situation, à mettre en relation les procédés à employer avec le but à atteindre, à choisir les moyens ou à découvrir les solutions originales qui permettent l'adaptation aux exigences de l'action. »
 - « l'intelligence est la capacité de résoudre par la pensée des problèmes nouveaux. »

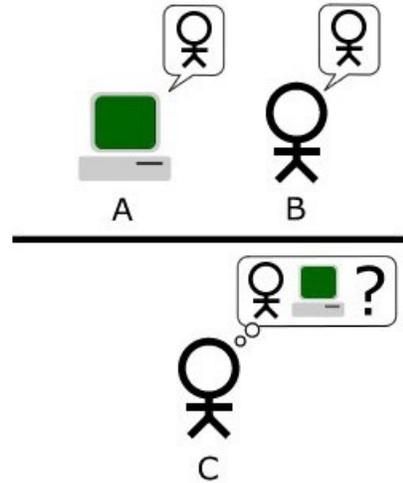
Pourquoi cet optimisme ?

- Difficile de définir « intelligence » !
- Quelques définitions :
 - « L'intelligence mesure la capacité d'un agent à atteindre ses objectifs dans un large éventail d'environnements. »
 - « [Dans des circonstances nouvelles pour lesquelles l'instinct, l'apprentissage passé ou l'habitude ne dispose d'aucune solution] Aptitude à appréhender et organiser les données de la situation, à mettre en relation les procédés à employer avec le but à atteindre, à choisir les moyens ou à découvrir les solutions originales qui permettent l'adaptation aux exigences de l'action. »
 - « l'intelligence est la capacité de résoudre par la pensée des problèmes nouveaux. »

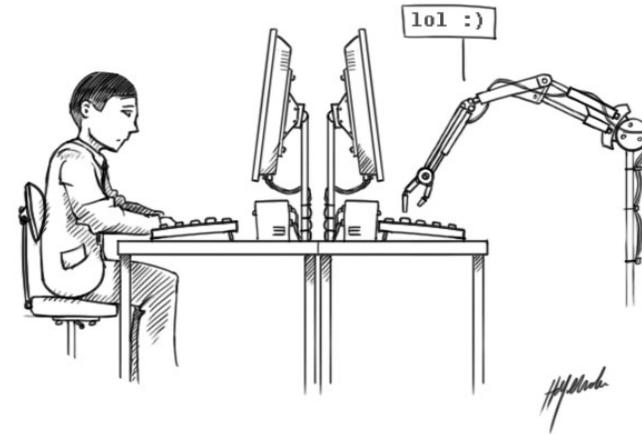
Comment la mesurer ?

Le jeu d'imitation/le test de Turing

Original :

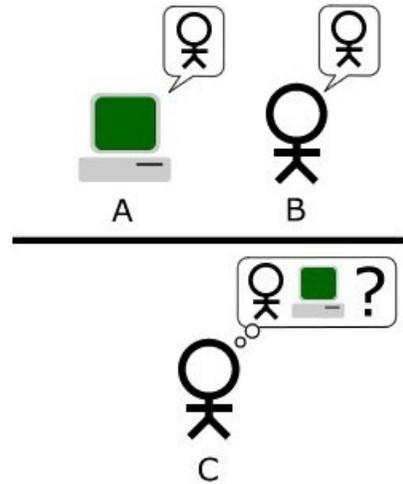


Interprétation courante:

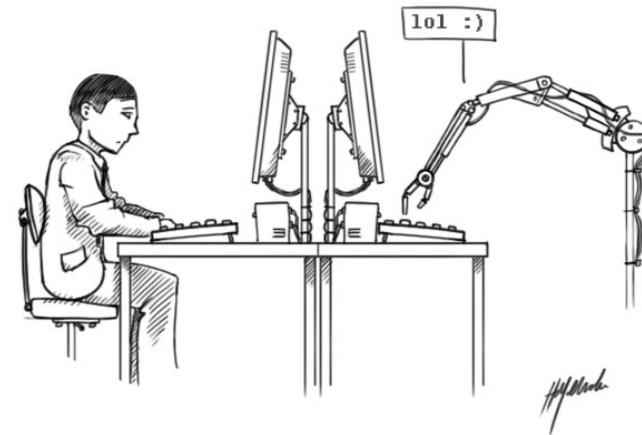


Le jeu d'imitation/le test de Turing

Original :



Interprétation courante:



On pourrait dire que ChatGPT et al. passent le test de Turing

Sauf : il y avait ELIZA

```
Welcome to
```

```
EEEEEE LL IIII ZZZZZZ AAAAA  
EE LL II ZZ AA AA  
EEEEEE LL II ZZZ AAAAAAA  
EE LL II ZZ AA AA  
EEEEEE LLLLLL IIII ZZZZZZ AA AA
```

```
Eliza is a mock Rogerian psychotherapist.  
The original program was described by Joseph Weizenbaum in 1966.  
This implementation by Norbert Landsteiner 2005.
```

```
ELIZA: Is something troubling you ?  
YOU: Men are all alike.  
ELIZA: What is the connection, do you suppose ?  
YOU: They're always bugging us about something or other.  
ELIZA: Can you think of a specific example ?  
YOU: Well, my boyfriend made me come here.  
ELIZA: Is it important to you that your boyfriend made you come here ?  
YOU: He says I'm depressed much of the time.  
ELIZA: I am sorry to hear that you are depressed.  
YOU: It's true. I am unhappy.  
ELIZA: Can you explain what made you unhappy ?  
YOU:
```

Sauf : il y avait ELIZA

Welcome to

```
EEEEEE LL IIII ZZZZZZ AAAAA  
EE LL II ZZ AA AA  
EEEEEE LL II ZZZ AAAAAA  
EE LL II ZZ AA AA  
EEEEEE LLLLLL IIII ZZZZZZ AA AA
```

Eliza is a mock Rogerian psychotherapist.

The original program was described by Joseph Weizenbaum in 1966.

This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?

YOU: Men are all alike.

ELIZA: What is the connection, do you suppose ?

YOU: They're always bugging us about something or other.

ELIZA: Can you think of a specific example ?

YOU: Well, my boyfriend made me come here.

ELIZA: Is it important to you that your boyfriend made you come here ?

YOU: He says I'm depressed much of the time.

ELIZA: I am sorry to hear that you are depressed.

YOU: It's true. I am unhappy.

ELIZA: Can you explain what made you unhappy ?

YOU:

Sauf : il y avait ELIZA

ELIZA : Quelque chose vous tracasse ?

VOUS : Les hommes sont tous pareils.

ELIZA : Quel est le lien, selon vous ?

VOUS : Ils nous embêtent toujours pour quelque chose ou autre.

ELIZA : Pouvez-vous penser à un exemple précis ?

VOUS : Eh bien, mon petit ami m'a fait venir ici.

ELIZA : Est-ce important pour vous que votre petit ami vous ait fait venir ici ?

VOUS : Il dit que je suis déprimée la plupart du temps.

ELIZA : Je suis désolée d'apprendre que vous êtes déprimée.

VOUS : C'est vrai. Je suis malheureuse.

ELIZA : Pouvez-vous expliquer ce qui vous a rendue malheureuse ?

Sauf : il y avait ELIZA

ELIZA : Quelque chose vous tracasse ?

VOUS : Les hommes sont tous pareils.

ELIZA : Quel est le lien, selon vous ?

VOUS : Ils nous embêtent toujours pour quelque chose ou autre.

ELIZA : Pouvez-vous penser à un exemple précis ?

VOUS : Eh bien, mon petit ami m'a fait venir ici.

ELIZA : Est-ce important pour vous que votre petit ami vous ait fait venir ici ?

VOUS : Il dit que je suis déprimée la plupart du temps.

ELIZA : Je suis désolée d'apprendre que vous êtes déprimée.

VOUS : C'est vrai. Je suis malheureuse.

ELIZA : Pouvez-vous expliquer ce qui vous a rendue malheureuse ?

Sauf : il y avait ELIZA

ELIZA : Quelque chose vous tracasse ?

VOUS : Les hommes sont tous pareils.

ELIZA : Quel est le lien, selon vous ?

VOUS : Ils nous embêtent toujours pour quelque chose ou autre.

ELIZA : Pouvez-vous penser à un exemple précis ?

VOUS : Eh bien, mon petit ami m'a fait venir ici.

ELIZA : Est-ce important pour vous que votre petit ami vous ait fait venir ici ?

VOUS : Il dit que je suis déprimée la plupart du temps.

ELIZA : Je suis désolée d'apprendre que vous êtes déprimée.

VOUS : C'est vrai. Je suis malheureuse.

ELIZA : Pouvez-vous expliquer ce qui vous a rendue malheureuse ?

Traduit par
Google Translate

Et aussi : PARRY

- Simulation d'une personne qui souffre de schizophrénie paranoïaque
- Utilisait dans une version du test de Turing en 1970 :
 - Vrais patients + PARRY
 - Deux groupes de psychiatres :
 - 1) Interaction directe
 - 2) Lecture de transcriptions
 - Qui ont correctement identifié le programme en **48 %** des cas = devinettes aléatoires

Pourquoi cet optimisme ? (2)

- Les chercheurs croient (et veulent croire) qu'une intelligence artificielle sera possible
 - Et promettre de développer l'intelligence artificielle permet d'avoir du financement
- Les entrepreneurs (Musk, Altman,...) ont un produit qui se vend mieux si le public croit qu'il s'agit de l'intelligence artificielle

Les succès de la recherche en « IA »

- Google Search
- Traduction automatique : Google Translate, DeepL
- Reconnaissance automatique de la parole : dicter un message
- Reconnaissance faciale : identifier les visages pour une photographie, déverrouiller le téléphone
- Reconnaissance de l'écriture manuscrite : lecture des chèques, lettres etc.
- Aide à la navigation : Waze, Google Maps
- Navigation autonome : Roomba
- Les systèmes de recommandation : Youtube, Spotify, Amazon etc.

- AlphaFold : prédiction de la structure des protéines

Les succès de la recherche en « IA »

- Google Search
- Traduction automatique : Google Translate, DeepL
- Reconnaissance automatique de la parole : dicter un message
- Reconnaissance faciale : identifier les visages pour une photographie, déverrouiller le téléphone
- Reconnaissance de l'écriture manuscrite : lecture des chèques, lettres etc.
- Aide à la navigation : Waze, Google Maps
- Navigation autonome : Roomba
- Les systèmes de recommandation : Youtube, Spotify, Amazon etc.
- AlphaFold : prédiction de la structure des protéines

**Prix Nobel de
chimie 2024**

Les succès de la recherche en « IA »

« Quand une technique rentre dans la vie quotidienne, on ne l'appelle plus 'IA' »

- Google Search
- Traduction automatique : Google Translate, DeepL
- Reconnaissance automatique de la parole : dicter un message
- Reconnaissance faciale : identifier les visages pour une photographie, déverrouiller le téléphone
- Reconnaissance de l'écriture manuscrite : lecture des chèques, lettres etc.
- Aide à la navigation : Waze, Google Maps
- Navigation autonome : Roomba
- Les systèmes de recommandation : Youtube, Spotify, Amazon etc.
- AlphaFold : prédiction de la structure des protéines

Prix Nobel de
chimie 2024

Mais qu'est-ce que c'est, l'« IA »

- Apprentissage
 - Supervisé
 - Non-supervisé
 - Par renforcement
- Représentation des connaissances
- Modélisation générative
- La robotique

Apprentissage supervisé

- On veut affecter une « **classe** » (y) à une entité
- On a une **description** (x)
- On suppose qu'il y a une relation dans le monde réel qui dérive la classe à partir de la description

Apprentissage supervisé

- On veut affecter une « **classe** » (y) à une entité
- On a une **description** (x)
- On suppose qu'il y a une relation dans le monde réel qui dérive la classe à partir de la description



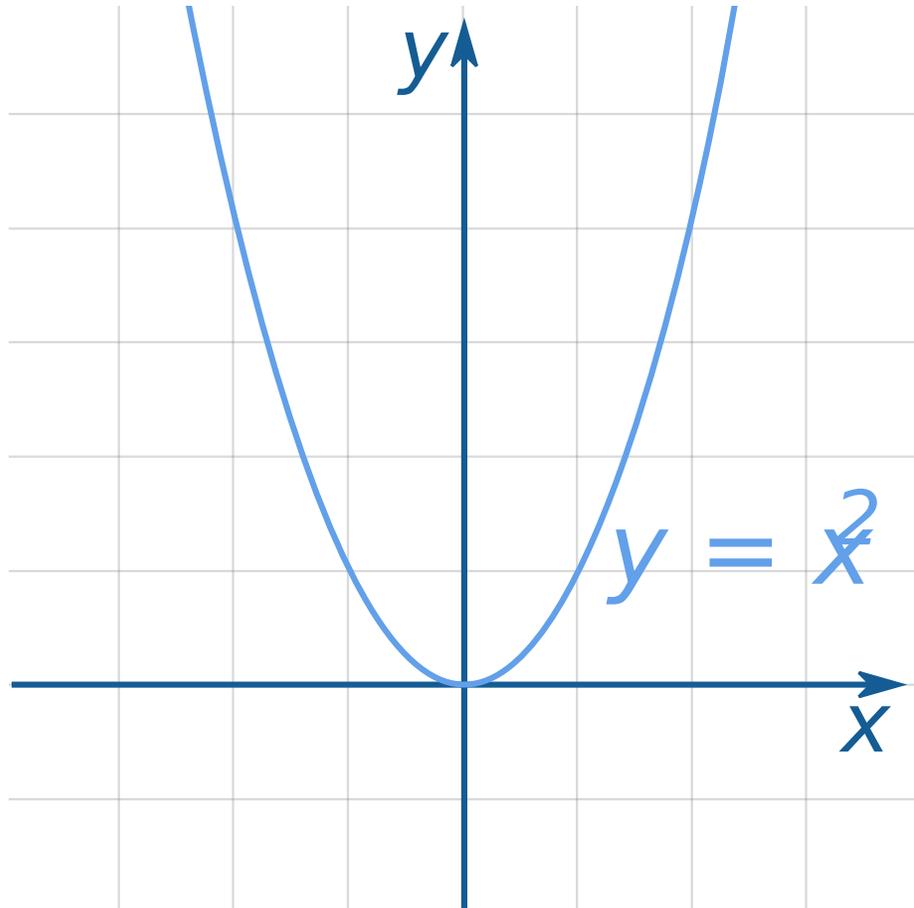
Apprentissage supervisé

- On veut affecter une « **classe** » (y) à une entité
- On a une **description** (x)
- On suppose qu'il y a une relation dans le monde réel qui dérive la classe à partir de la description

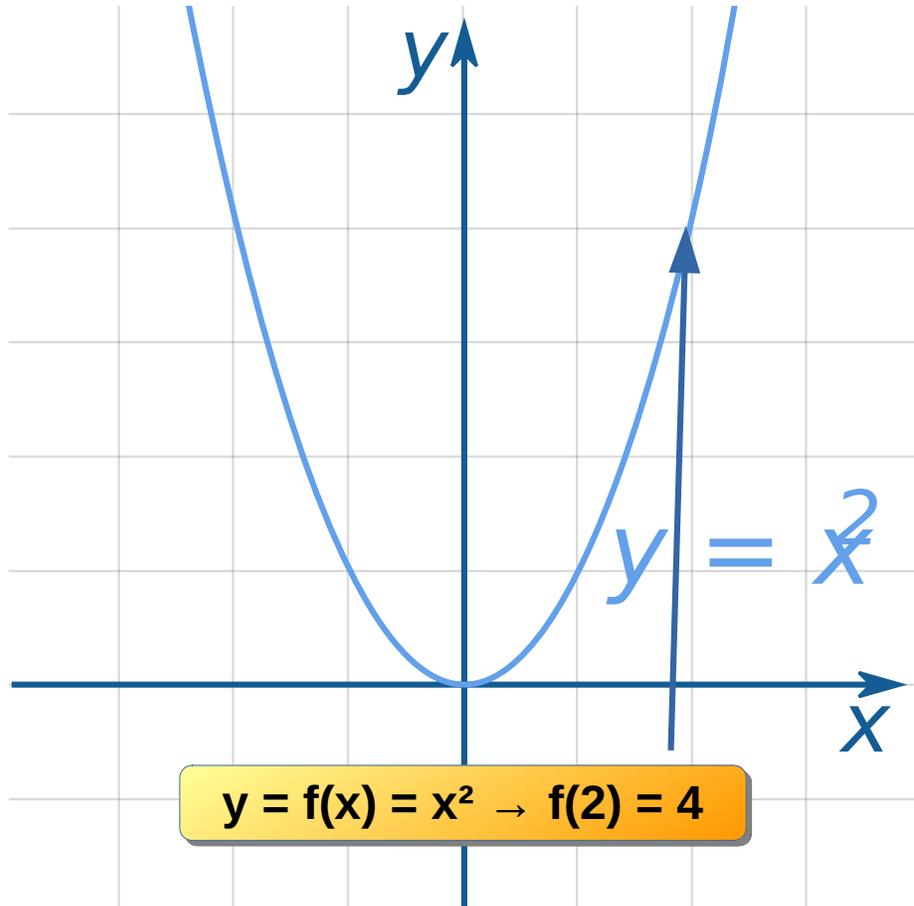
Coccinella septempunctata



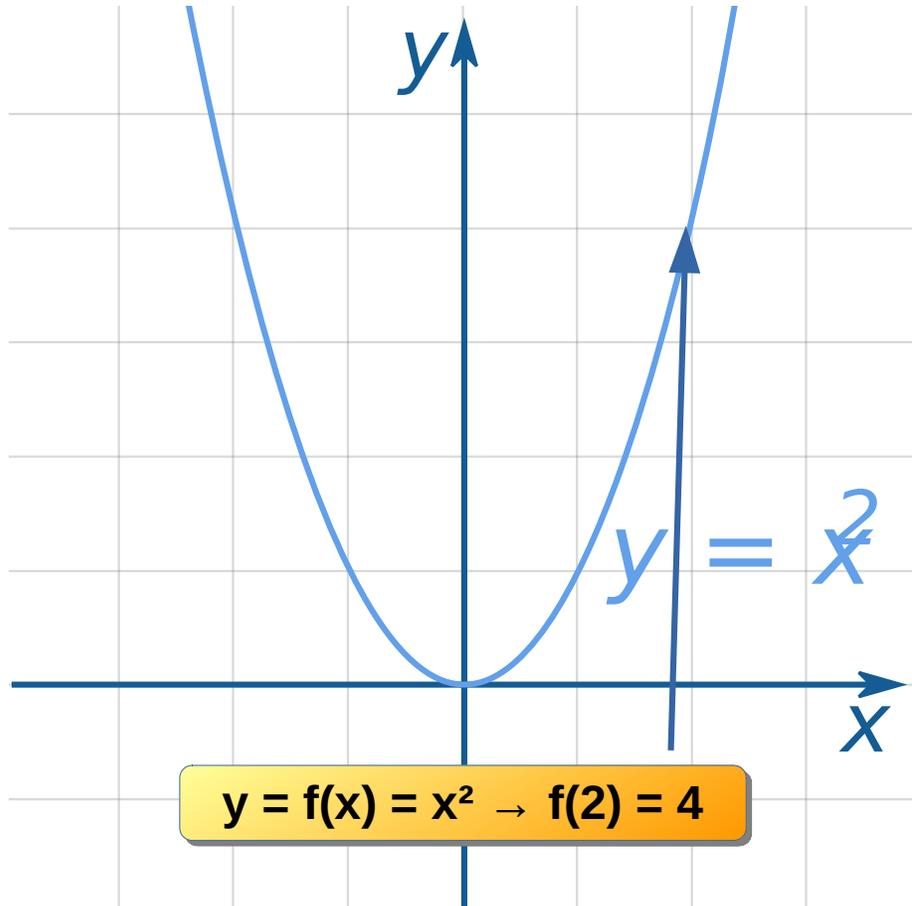
Fonction d'étiquetage



Fonction d'étiquetage

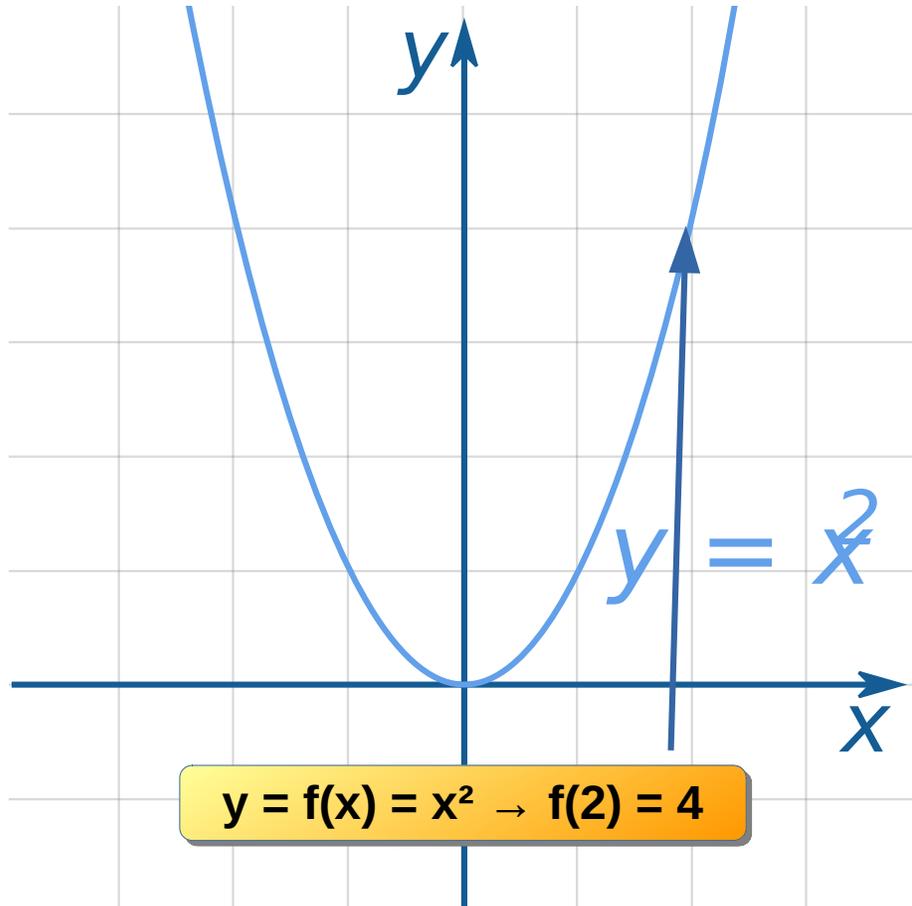


Fonction d'étiquetage



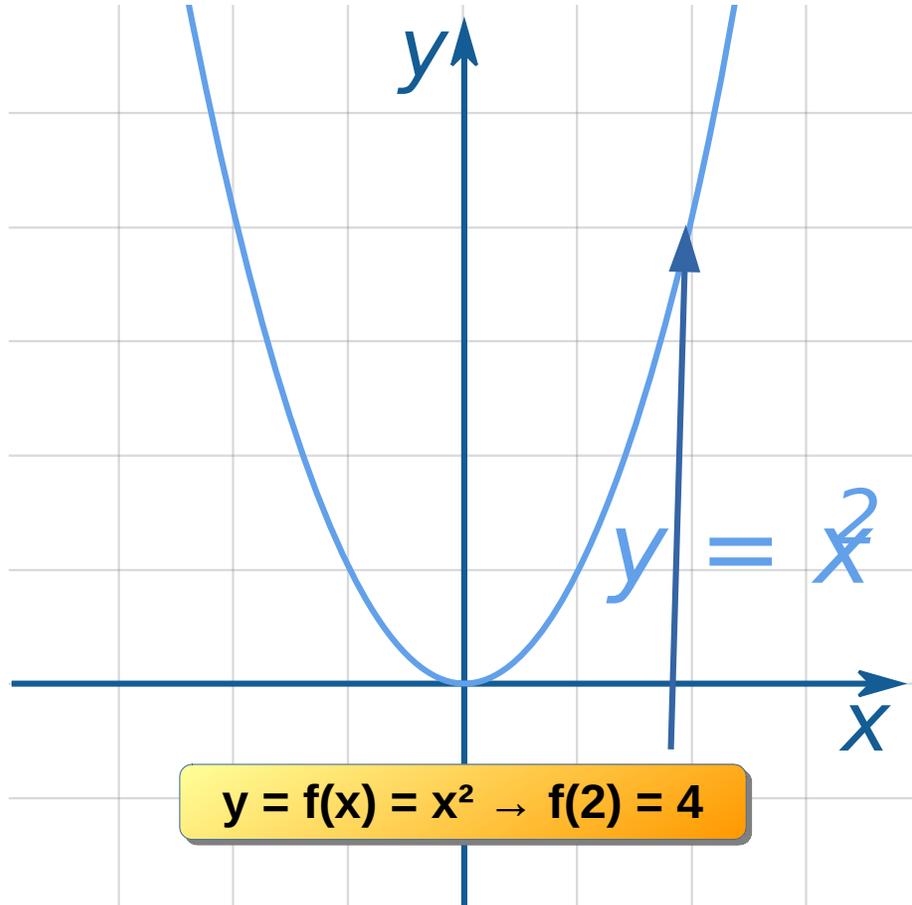
f() = Coccinella
septempunctata

Fonction d'étiquetage



- f() = Coccinella septempunctata
- On ne connaît pas la forme de la fonction

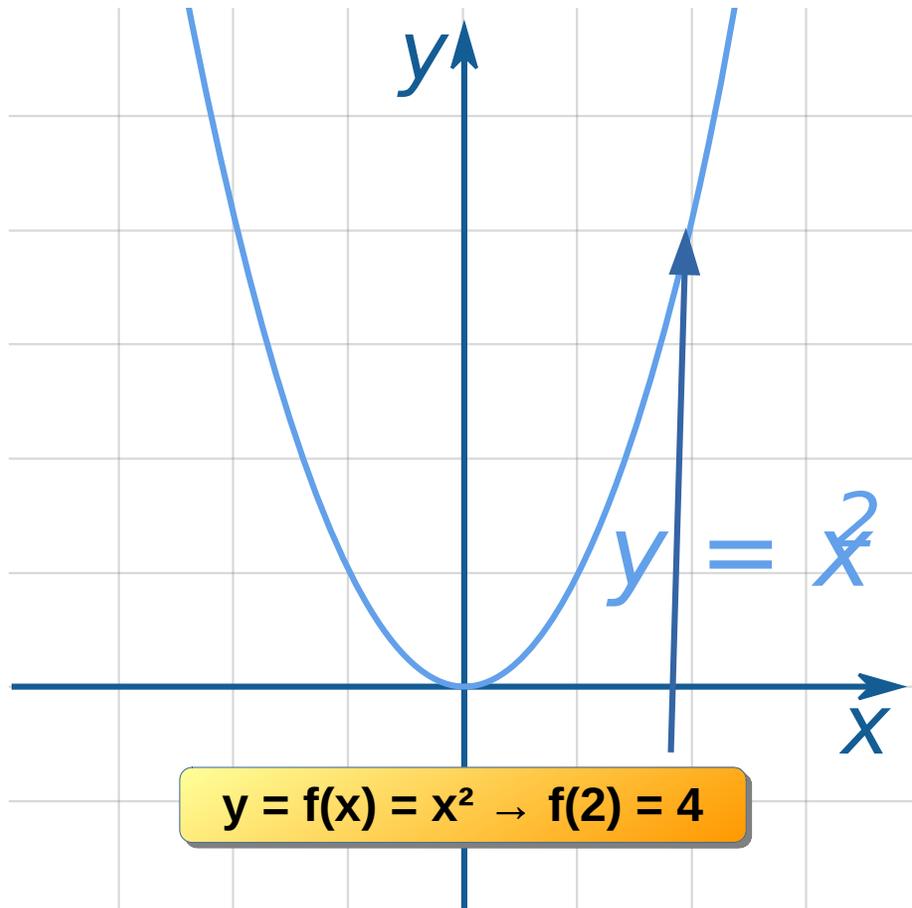
Fonction d'étiquetage



$f(\text{img alt="Coccinella septempunctata" data-bbox="538 218 666 378}}) = \text{Coccinella septempunctata}$

- On ne connaît pas la forme de la fonction
- Solution :
 - 1) On prend beaucoup d'exemples d'entités avec leurs étiquettes
 - 2) On choisit un type de modèle
 - 3) On apprend une approximation f' de la fonction

Fonction d'étiquetage



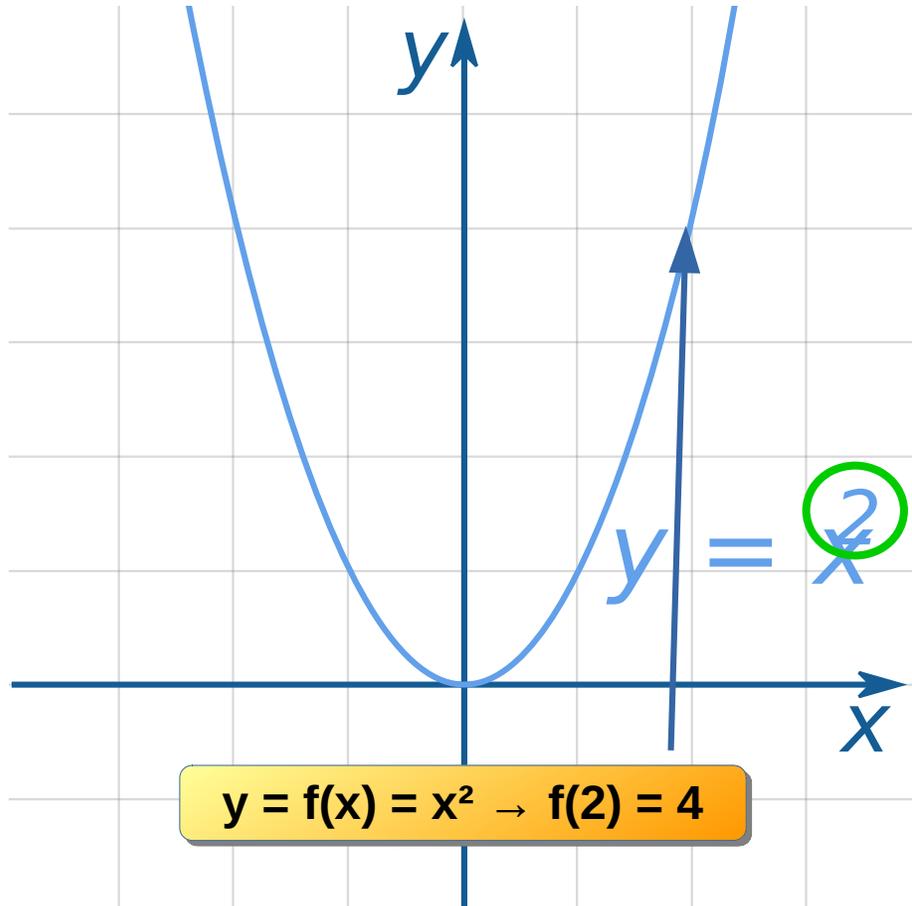
f() = Coccinella septempunctata

- On ne connaît pas la forme de la fonction

• Solution :

- 1) On prend b  avec leurs é  ex  n 
- 2) On choisit u  nc 
- 3) On apprend une approxi  le la  fonction

Fonction d'étiquetage



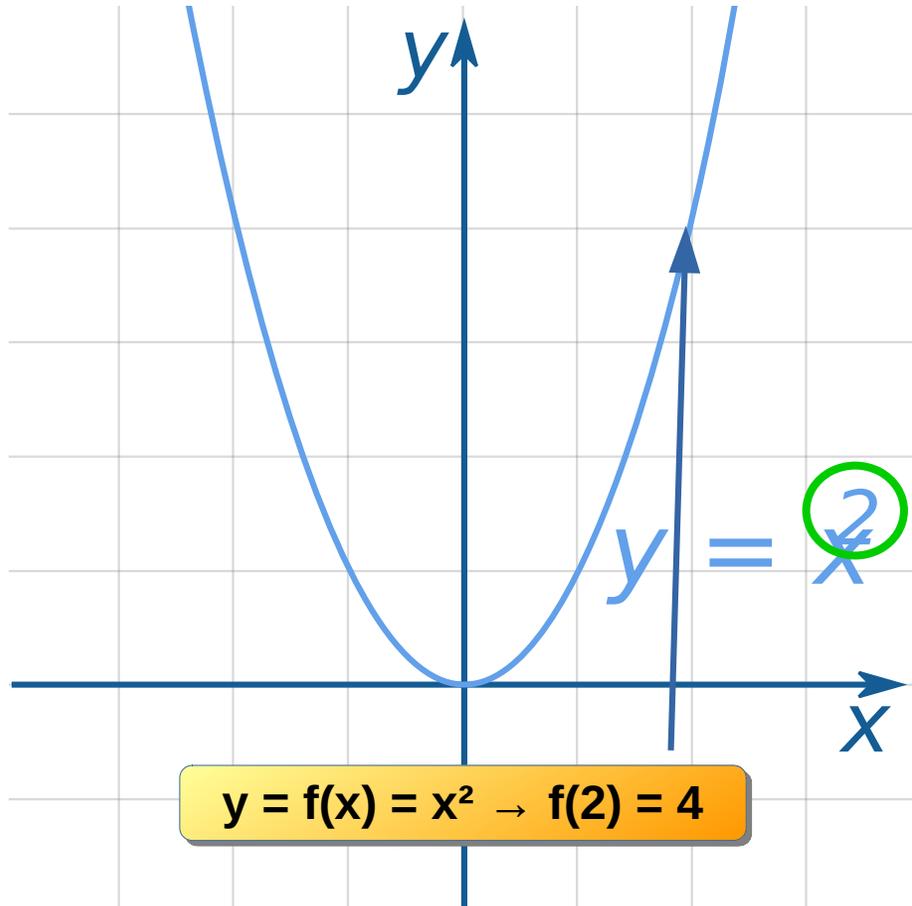
f() = Coccinella septempunctata

- On ne connaît pas la forme de la fonction

• Solution :

- 1) On prend b  avec leurs é  ex  n 
- 2) On choisit u  nc 
- 3) On apprend une approxi  le la  fonction 

Fonction d'étiquetage



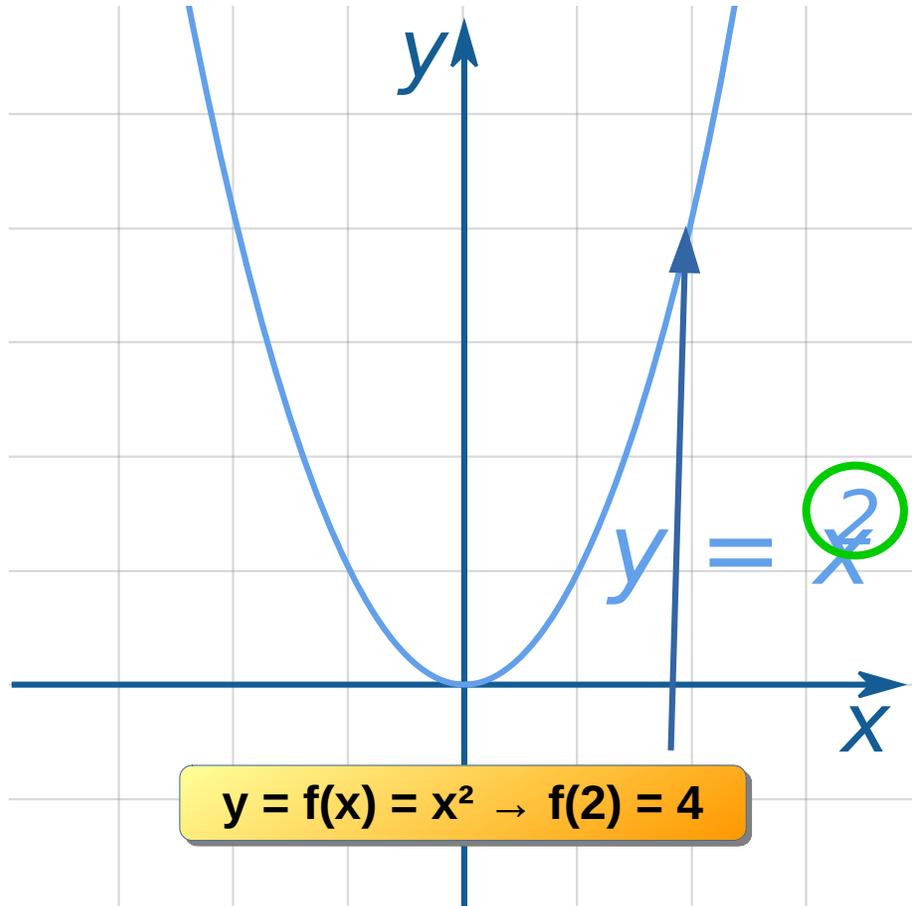
f() = Coccinella septempunctata

- On ne connaît pas la forme de la fonction

- Solution :

- 1) On prend b  avec leurs é  ex  n 
- 2) On choisit u  nc 
- 3) On apprend une approxi  le la  fonction 

Fonction d'étiquetage

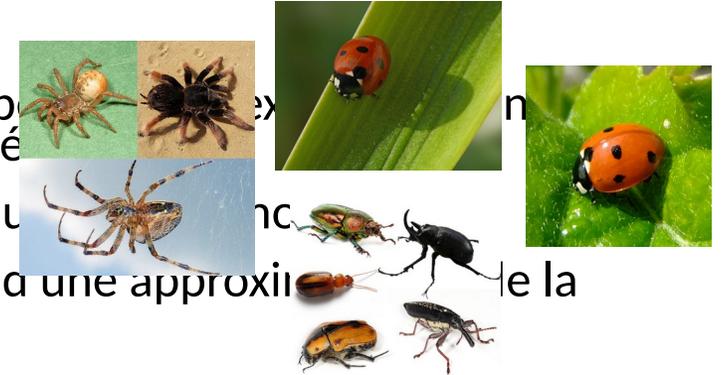


$f(\text{img alt="Coccinella septempunctata" data-bbox="538 218 665 375}}) = \text{Coccinella septempunctata}$

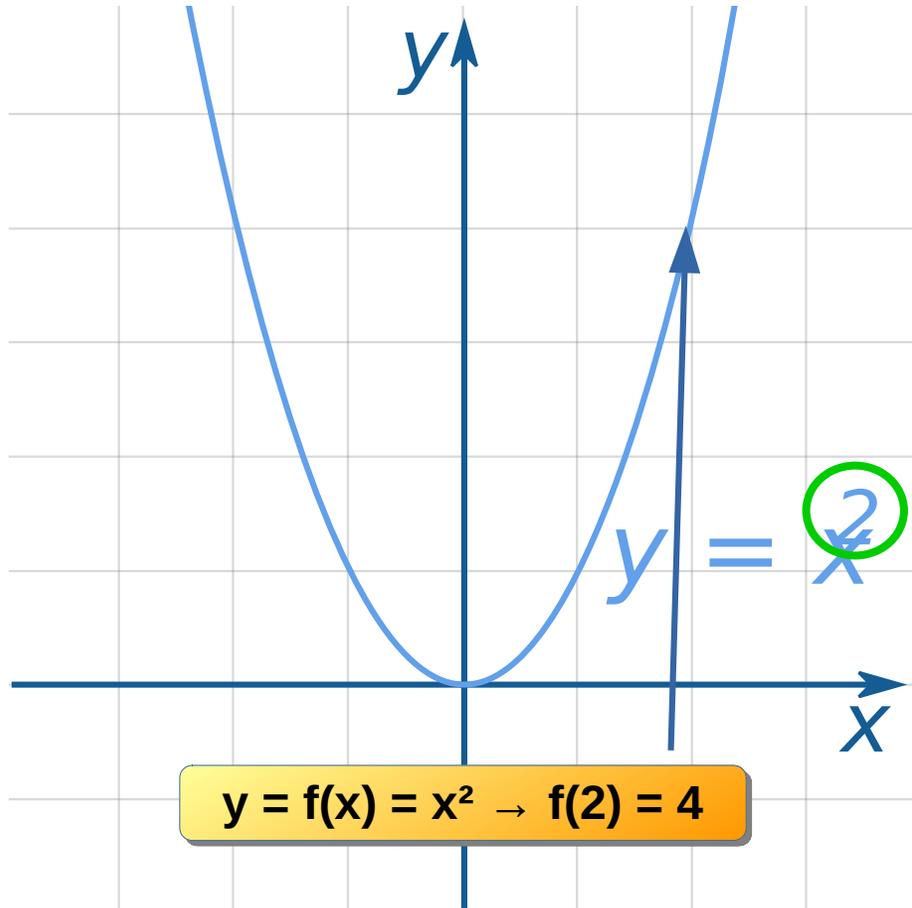
- On ne connaît pas la forme de la fonction

- Solution :

- 1) On prend b...
avec leurs é...
- 2) On choisit u...
nc...
- 3) On apprend une approxi...
le la fonction



Fonction d'étiquetage



f() = Coccinella septempunctata

- On ne connaît pas la forme de la fonction

- Solution :

- 1) On prend b  et on les étiquette avec leurs espèces  et  n
- 2) On choisit un  et on les étiquette avec leurs espèces  n
- 3) On apprend une approximation de la fonction  et  n

- P.ex. : si  et  → C.s.

Apprentissage supervisé : quelques exemples

- Prédiction du remboursement d'un prêt
 - X : apport, travail, biens
 - Y : le prêt sera remboursé/non-remboursé
- Prédiction de la présence d'un cancer
 - X : Radiographie
 - Y : s'il y a ou pas un cancer
- Prédiction d'une activité thérapeutique
 - X : représentation d'une molécule (atomes et liaisons, descripteurs chimiques)
 - Y : activité par rapport à une cible thérapeutique
- Prédiction du victoire d'un match sportif
 - X : statistiques de performances des équipes
 - Y : victoire/défaites de l'équipe à domicile

Apprentissage supervisé : quelques exemples

- Prédiction du remboursement d'un prêt
 - X : apport, travail, biens
 - Y : le prêt sera remboursé/non-remboursé
- Prédiction de la présence d'un cancer
 - X : Radiographie
 - Y : s'il y a ou pas un cancer
- Prédiction d'une activité thérapeutique
 - X : représentation d'une molécule (atomes et liaisons, descripteurs chimiques)
 - Y : activité par rapport à une cible thérapeutique
- Prédiction du victoire d'un match sportif
 - X : statistiques de performances des équipes
 - Y : victoire/défaites de l'équipe à domicile

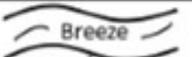
Aussi :

- Utiliser Dall-E pour créer une image spécifique
- Utiliser Mistral pour créer du texte dans le style de Hemingway

Apprentissage par renforcement

- On veut arriver à un but (naviguer vers un endroit, manipuler un objet, gagner au Blackjack)
 - Mais on ne sait pas vraiment comment
- On peut décrire l'état actuel dans lequel on se trouve
 - Il y a des états finaux dont un (ou plusieurs) représentent le but
- On a une certaine nombre d'actions qu'on peut utiliser

Exemple : Le monde du Wumpus

4	 Stench		 Breeze	
3	 Wumpus	 Breeze  Stench  Gold		 Breeze
2	 Stench		 Breeze	
1	 Agent	 Breeze		 Breeze
	1	2	3	4

À côté du Wumpus : odeur nauséabonde

À côté d'une fosse : brise

En présence du trésor : ça brille

On marche dans un mur : choc

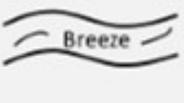
Le Wumpus meurt : il cri (et on gagne 1000)

État : {Coordonnées, ON, Brise, ÇB, Choc, Cri, Orientation}

Arriver à la case (1,1) avec le trésor : **1000**

Tuer le Wumpus : **100**

Exemple : Le monde du Wumpus

4	 Stench		 Breeze	
3	 Wumpus	 Breeze  Stench  Gold		 Breeze
2	 Stench		 Breeze	
1	 Agent	 Breeze		 Breeze
	1	2	3	4

Arriver à la case (1,1) avec le trésor : **1000**

Tuer le Wumpus : **100**

À côté du Wumpus : odeur nauséabonde

À côté d'une fosse : brise

En présence du trésor : ça brille

On marche dans un mur : choc

Le Wumpus meurt : il cri (et on gagne 1000)

État : {Coordonnées, ON, Brise, ÇB, Choc, Cri, Orientation}

Actions possible :

Tourner gauche/droite

Avancer

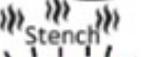
Prendre le trésor

Lâcher le trésor

Tirer une flèche

Besoin de deux mains pour tirer

Exemple : Le monde du Wumpus

4	 Stench		 Breeze	
3	 Wumpus	 Breeze	 Stench	 Gold
2	 Stench		 Breeze	
1	 Agent	 Breeze	 Breeze	
	1	2	3	4

Arriver à la case (1,1) avec le trésor : 

Tuer le Wumpus : 

À côté du Wumpus : odeur nauséabonde

À côté d'une fosse : brise

En présence du trésor : ça brille

On marche dans un mur : choc

Le Wumpus meurt : il cri (et on gagne 1000)

État : {Coordonnées, ON, Brise, ÇB, Choc, Cri, Orientation}

Actions possible :

Tourner gauche/droite

Avancer

Prendre le trésor

Lâcher le trésor

Tirer une flèche

Besoin de deux mains pour tirer

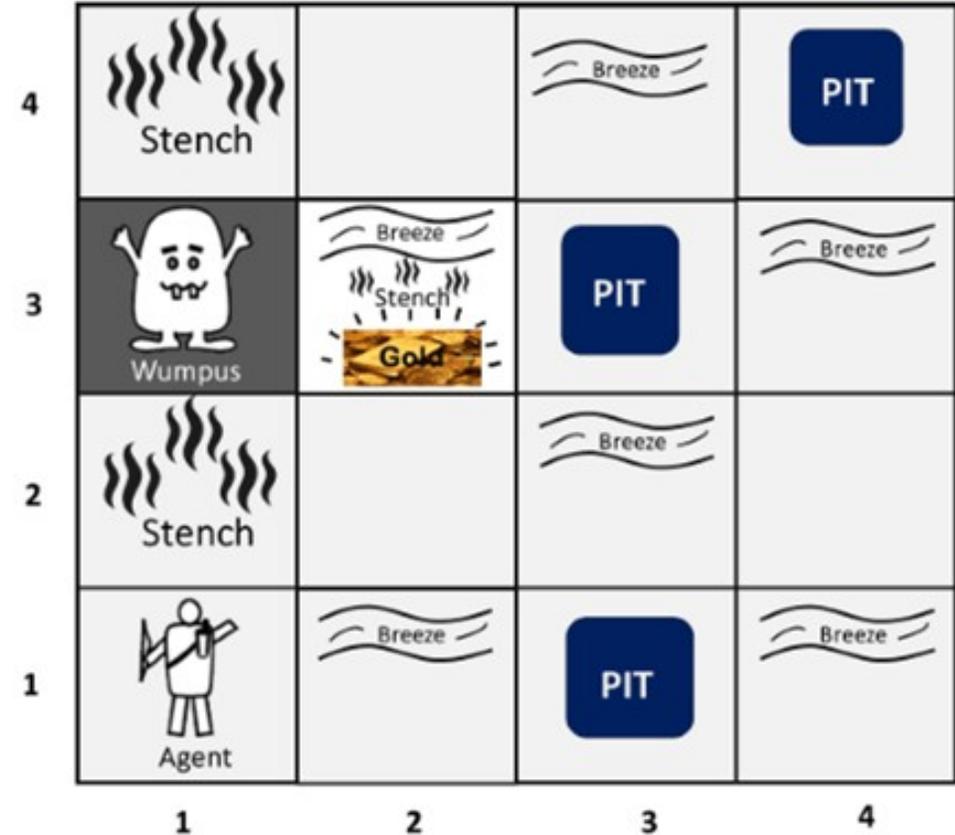
Apprentissage par renforcement (2)

- Faire une action
 - Au début aléatoirement
 - Plus tard la meilleure action
- Passer d'un état à un autre
- Quand on arrive à un état final, affecter la récompense/la punition aux combinaisons **état/action** précédentes
 - La dernière combinaison reçoit 100 %
 - L'avant-dernière 90 % etc.
- Résultat : une liste de combinaisons état/action avec des valeurs affectées
 - Dans chaque état on peut choisir l'action avec la meilleure valeur

4	 Stench		 Breeze	
3	 Wumpus	 Breeze  Stench  Gold		
2	 Stench			
1	 Agent			
	1	2	3	4

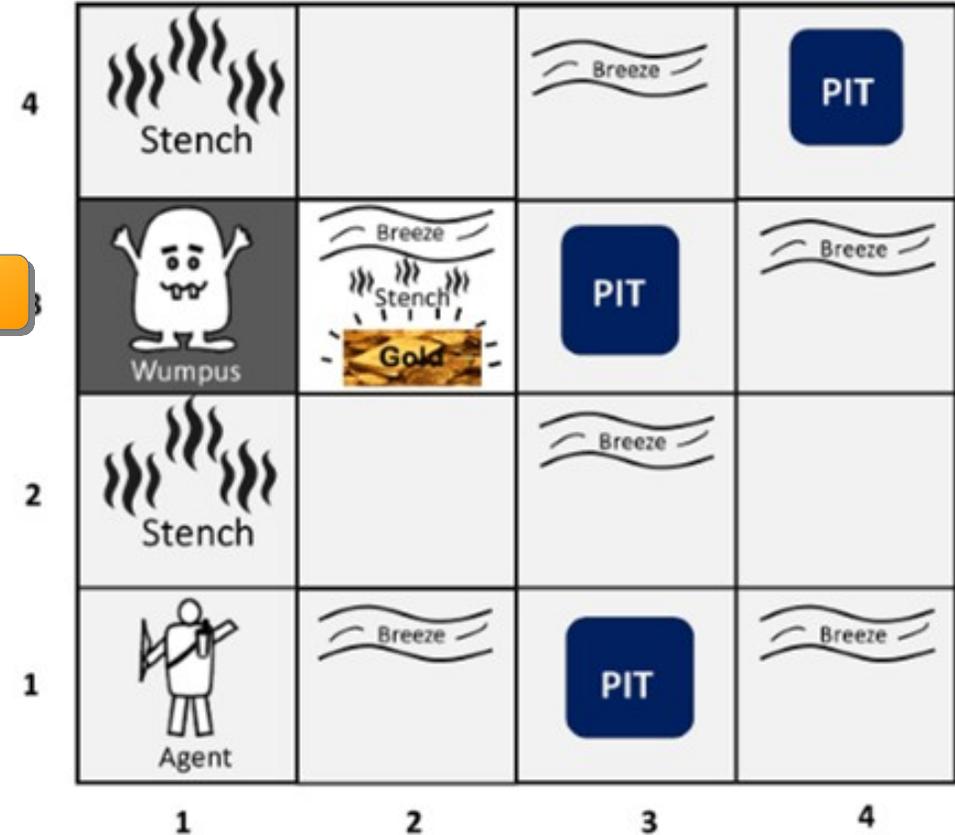
Apprentissage par renforcement (2)

- Faire une action **Par ex. se déplacer vers le nord**
 - Au début aléatoirement
 - Plus tard la meilleure action
- Passer d'un état à un autre
- Quand on arrive à un état final, affecter la récompense/la punition aux combinaisons **état/action** précédentes
 - La dernière combinaison reçoit 100 %
 - L'avant-dernière 90 % etc.
- Résultat : une liste de combinaisons état/action avec des valeurs affectées
 - Dans chaque état on peut choisir l'action avec la meilleure valeur



Apprentissage par renforcement (2)

- Faire une action **Par ex. se déplacer vers le nord**
 - Au début aléatoirement
 - Plus tard la meilleure action
- Passer d'un état à un autre **P.ex. (1,1) → (2,1)**
- Quand on arrive à un état final, affecter la récompense/la punition aux combinaisons **état/action** précédentes
 - La dernière combinaison reçoit 100 %
 - L'avant-dernière 90 % etc.
- Résultat : une liste de combinaisons état/action avec des valeurs affectées
 - Dans chaque état on peut choisir l'action avec la meilleure valeur



Apprentissage par renforcement (2)

- Faire une action **Par ex. se déplacer vers le nord**
 - Au début aléatoirement
 - Plus tard la meilleure action
- Passer d'un état à un autre **P.ex. (1,1) → (2,1)**
- Quand on arrive à un état final, affecter la récompense/la punition aux combinaisons **état/action** précédentes
 - La dernière combinaison reçoit 100 %
 - L'avant-dernière 90 % etc.
- Résultat : une liste de combinaisons état/action avec des valeurs affectées
 - Dans chaque état on peut choisir l'action avec la meilleure valeur

(2,1), nord = -1000
(1,1), nord = -900

4	 Stench		 Breeze	 PIT
3	 Wumpus	 Breeze  Stench  Gold	 PIT	 Breeze
2	 Stench		 Breeze	
1	 Agent	 Breeze	 PIT	 Breeze
	1	2	3	4

Représentation des connaissances

On veut rendre accessible des connaissances pour l'utilisation par des programmes, p.ex. : **graphe de connaissances**

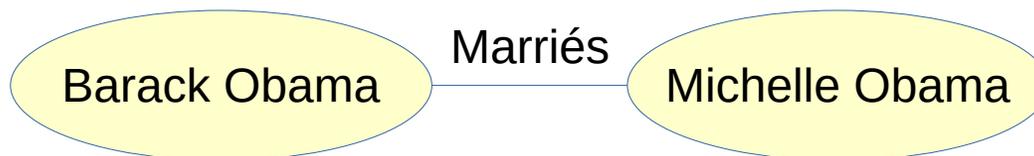


Barack Obama

Michelle Obama

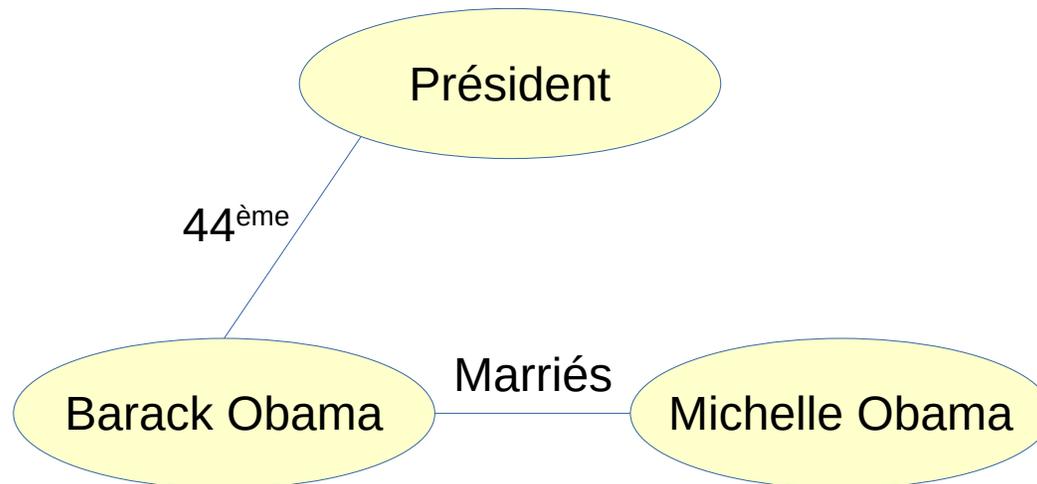
Représentation des connaissances

On veut rendre accessible des connaissances pour l'utilisation par des programmes, p.ex. : **graphe de connaissances**



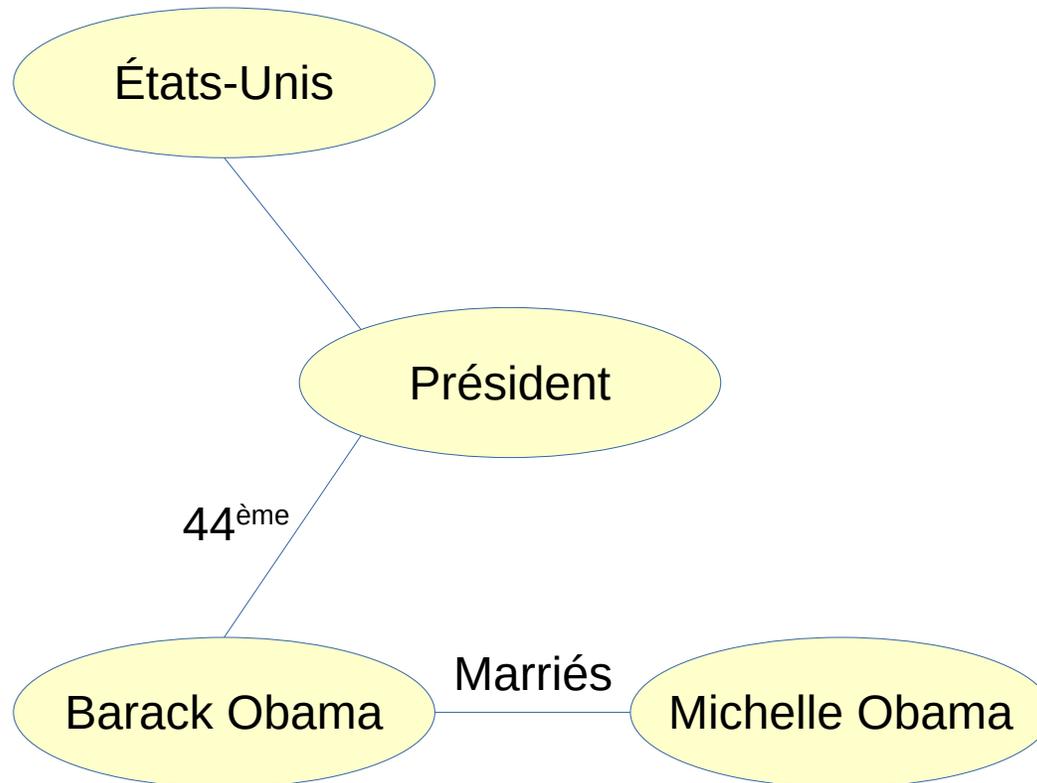
Représentation des connaissances

On veut rendre accessible des connaissances pour l'utilisation par des programmes, p.ex. : **graphe de connaissances**



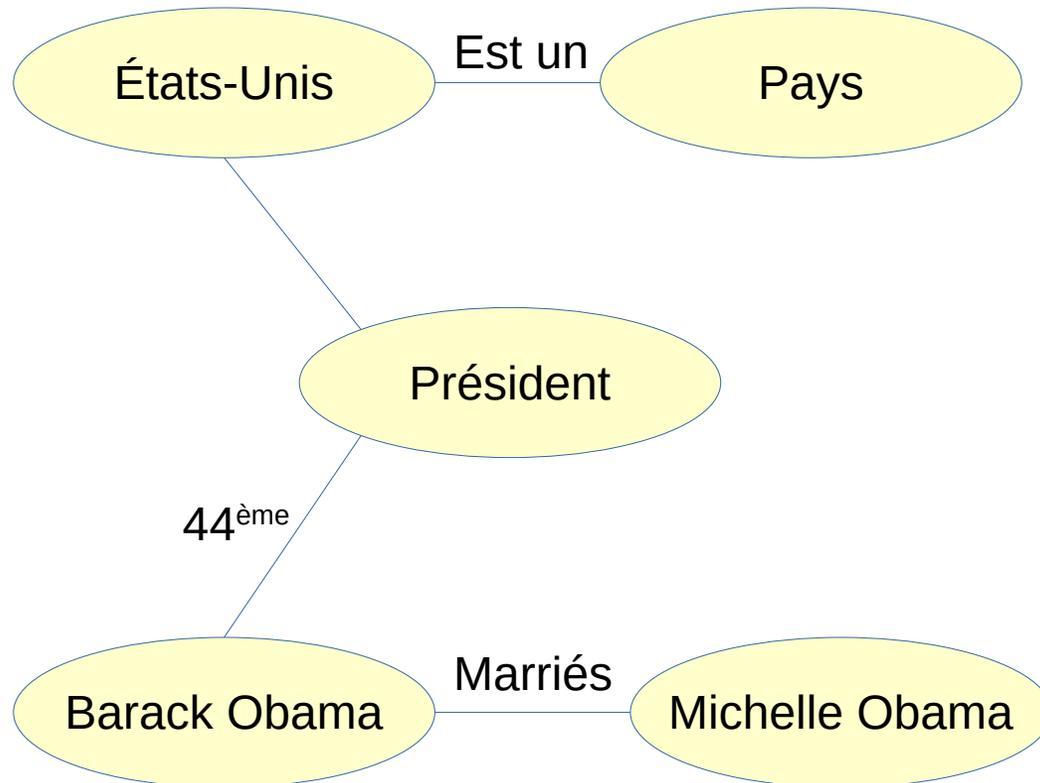
Représentation des connaissances

On veut rendre accessible des connaissances pour l'utilisation par des programmes, p.ex. : **graphe de connaissances**



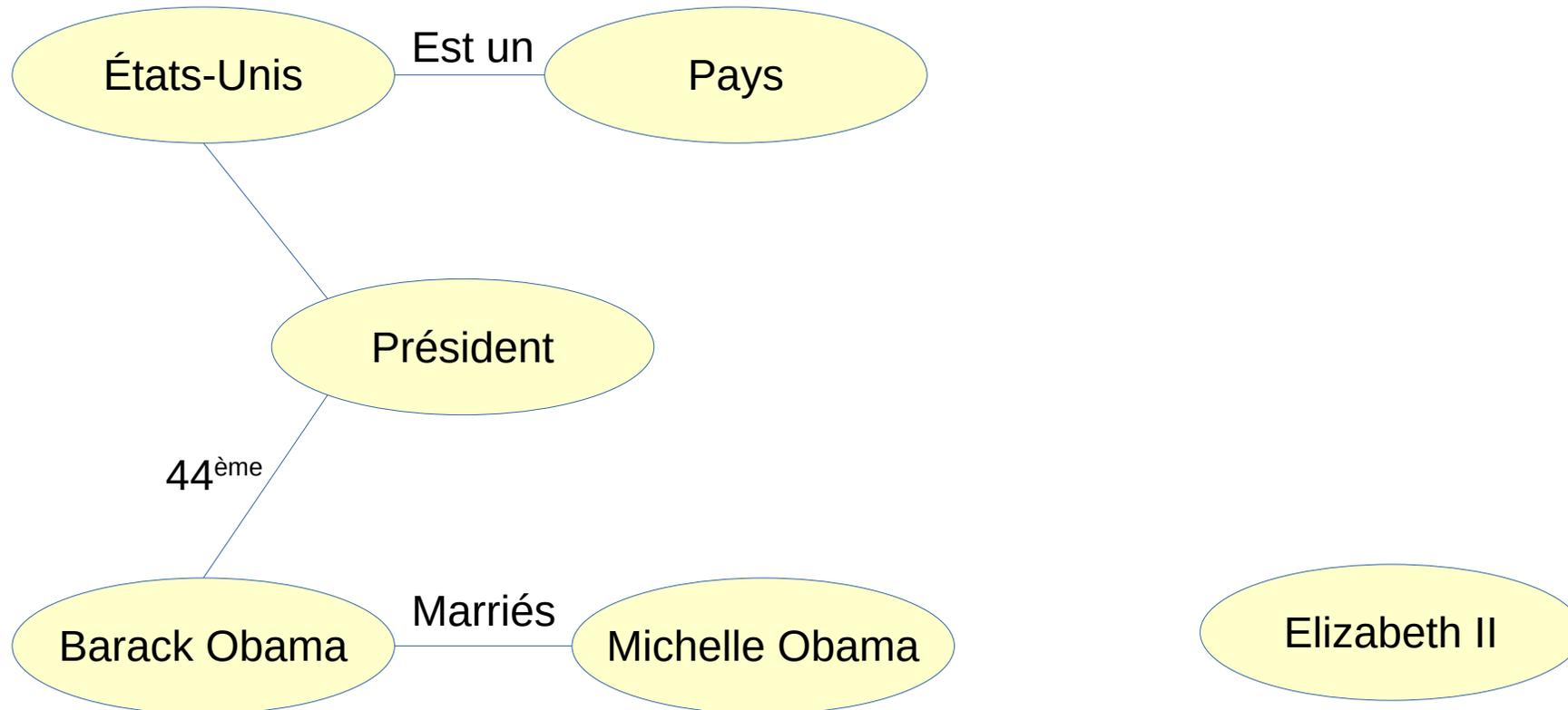
Représentation des connaissances

On veut rendre accessible des connaissances pour l'utilisation par des programmes, p.ex. : **graphe de connaissances**



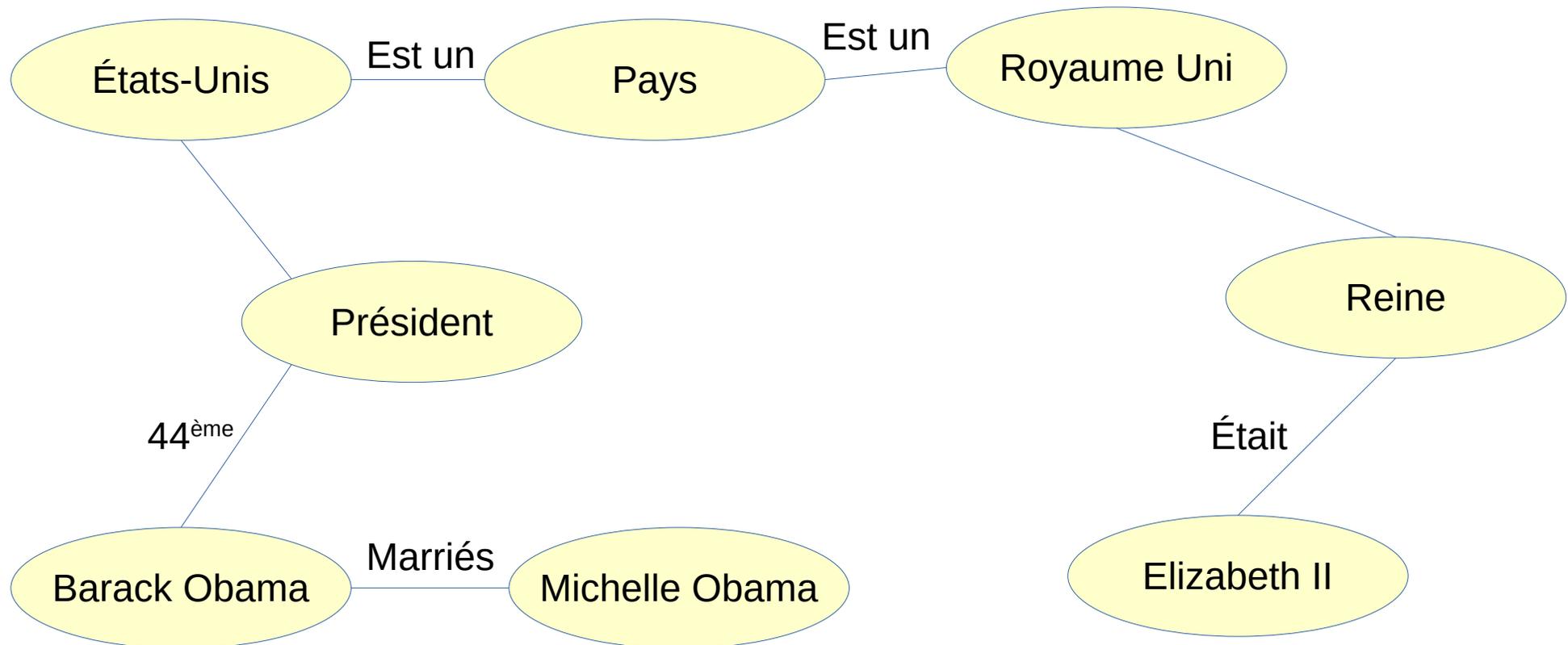
Représentation des connaissances

On veut rendre accessible des connaissances pour l'utilisation par des programmes, p.ex. : **graphe de connaissances**



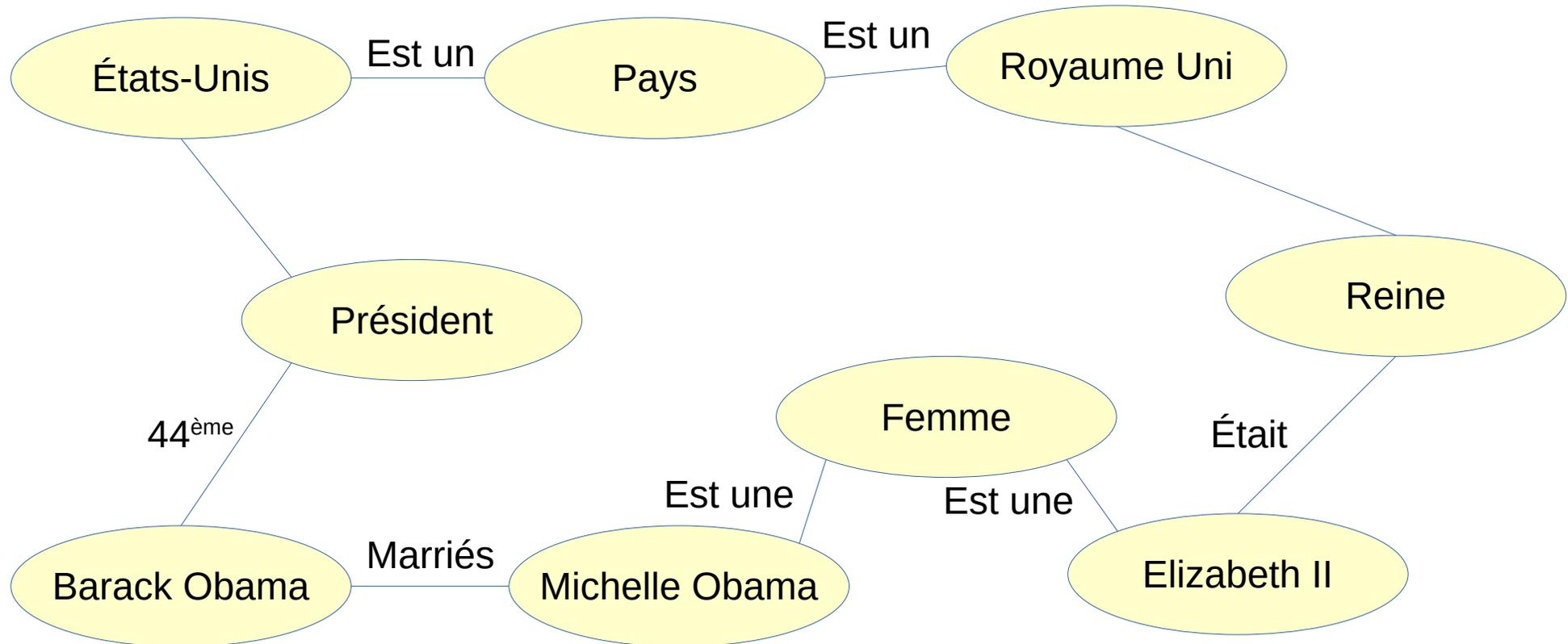
Représentation des connaissances

On veut rendre accessible des connaissances pour l'utilisation par des programmes, p.ex. : **graphe de connaissances**



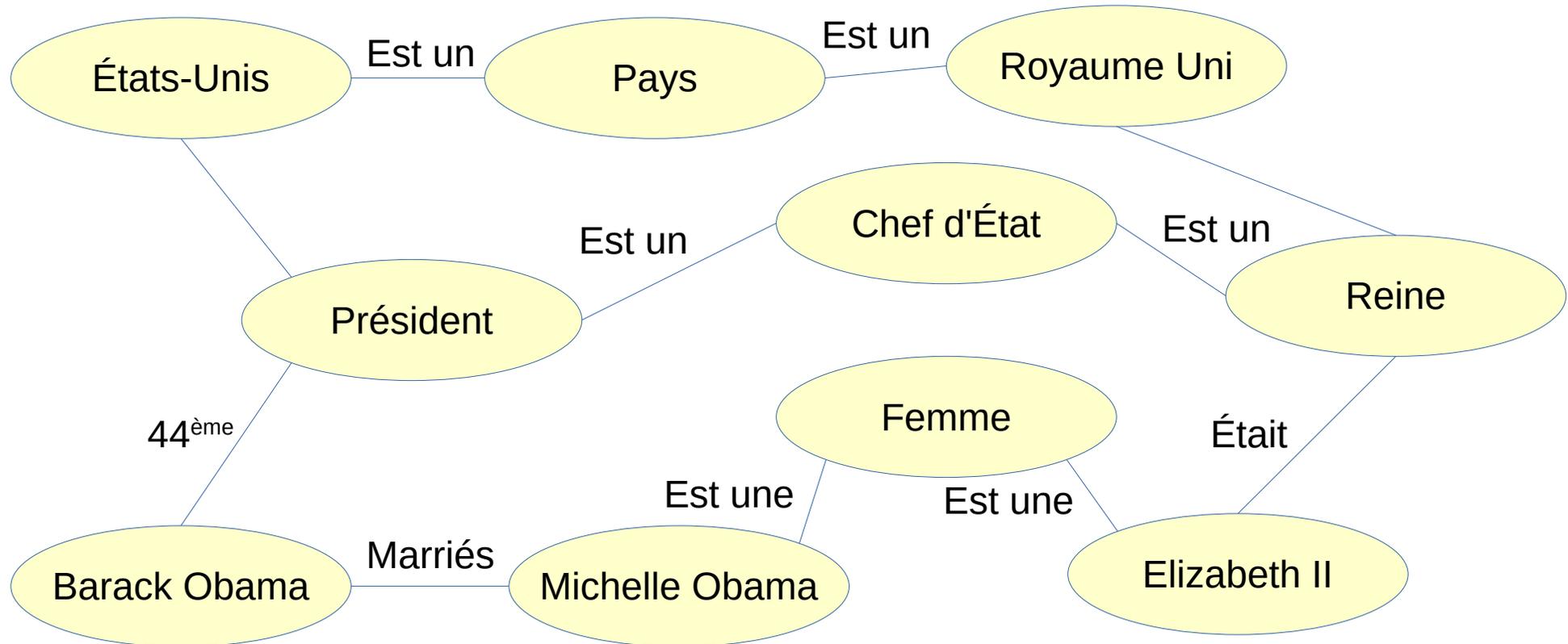
Représentation des connaissances

On veut rendre accessible des connaissances pour l'utilisation par des programmes, p.ex. : **graphe de connaissances**



Représentation des connaissances

On veut rendre accessible des connaissances pour l'utilisation par des programmes, p.ex. : **graphe de connaissances**



Modélisation générative

ChatGPT et al. sont
des modèles génératifs

- On veut, à partir d'un contexte générer un « token »
 - Étant donné une phrase partielle, quel est le prochain mot ?
 - Étant donné des pixels de différentes couleurs, quelle est la couleur du pixel à côté ?
- On apprend des **probabilités** étant donné le contexte

Le

Modélisation générative

ChatGPT et al. sont
des modèles génératifs

- On veut, à partir d'un contexte générer un « token »
 - Étant donné une phrase partielle, quel est le prochain mot ?
 - Étant donné des pixels de différentes couleurs, quelle est la couleur du pixel à côté ?
- On apprend des **probabilités** étant donné le contexte

Prompt pour Gemini
Image pour Midjourney

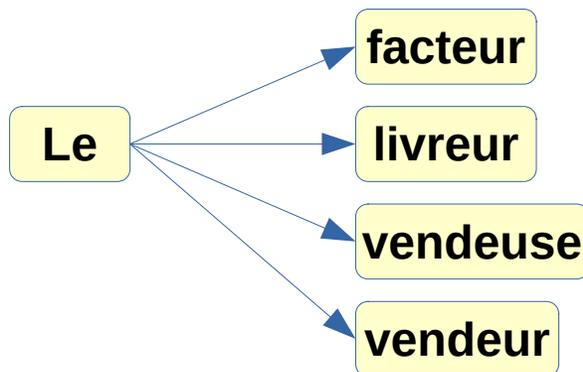
Le

Modélisation générative

ChatGPT et al. sont
des modèles génératifs

- On veut, à partir d'un contexte générer un « token »
 - Étant donné une phrase partielle, quel est le prochain mot ?
 - Étant donné des pixels de différentes couleurs, quelle est la couleur du pixel à coté ?
- On apprend des **probabilités** étant donné le contexte

Prompt pour Gemini
Image pour Midjourney

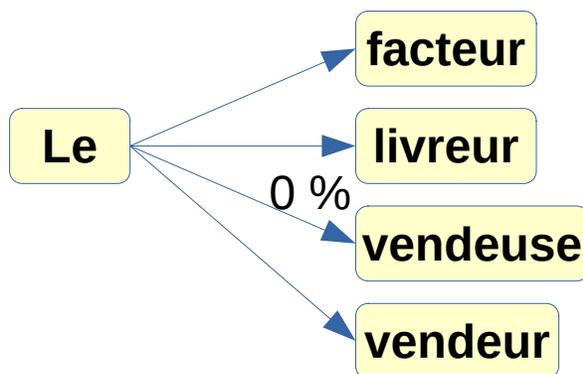


Modélisation générative

ChatGPT et al. sont
des modèles génératifs

- On veut, à partir d'un contexte générer un « token »
 - Étant donné une phrase partielle, quel est le prochain mot ?
 - Étant donné des pixels de différentes couleurs, quelle est la couleur du pixel à côté ?
- On apprend des **probabilités** étant donné le contexte

Prompt pour Gemini
Image pour Midjourney

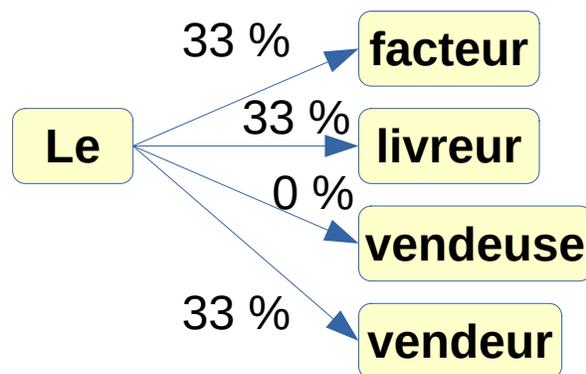


Modélisation générative

ChatGPT et al. sont des modèles génératifs

- On veut, à partir d'un contexte générer un « token »
 - Étant donné une phrase partielle, quel est le prochain mot ?
 - Étant donné des pixels de différentes couleurs, quelle est la couleur du pixel à coté ?
- On apprend des **probabilités** étant donné le contexte

Prompt pour Gemini
Image pour Midjourney



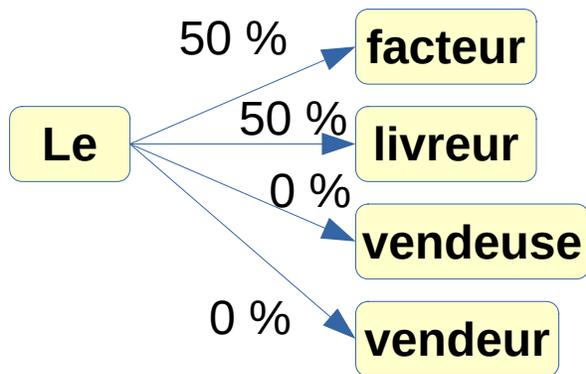
Modélisation générative

ChatGPT et al. sont des modèles génératifs

- On veut, à partir d'un contexte générer un « token »
 - Étant donné une phrase partielle, quel est le prochain mot ?
 - Étant donné des pixels de différentes couleurs, quelle est la couleur du pixel à coté ?
- On apprend des **probabilités** étant donné le contexte

Prompt pour Gemini
Image pour Midjourney

La camionnette est arrivée à 9 heures du matin.



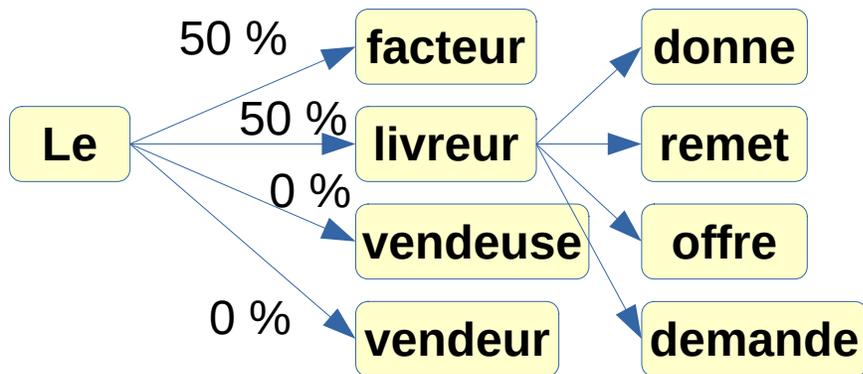
Modélisation générative

ChatGPT et al. sont des modèles génératifs

- On veut, à partir d'un contexte générer un « token »
 - Étant donné une phrase partielle, quel est le prochain mot ?
 - Étant donné des pixels de différentes couleurs, quelle est la couleur du pixel à coté ?
- On apprend des **probabilités** étant donné le contexte

Prompt pour Gemini
Image pour Midjourney

La camionnette est arrivée à 9 heures du matin.



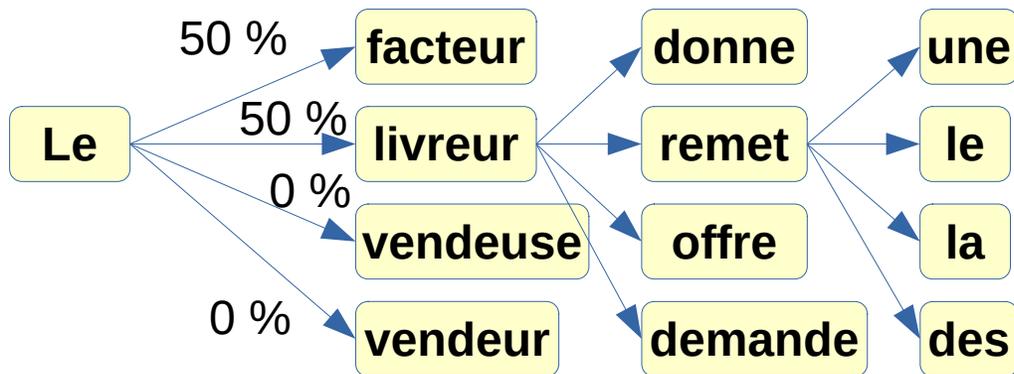
Modélisation générative

ChatGPT et al. sont des modèles génératifs

- On veut, à partir d'un contexte générer un « token »
 - Étant donné une phrase partielle, quel est le prochain mot ?
 - Étant donné des pixels de différentes couleurs, quelle est la couleur du pixel à coté ?
- On apprend des **probabilités** étant donné le contexte

Prompt pour Gemini
Image pour Midjourney

La camionnette est arrivée à 9 heures du matin.



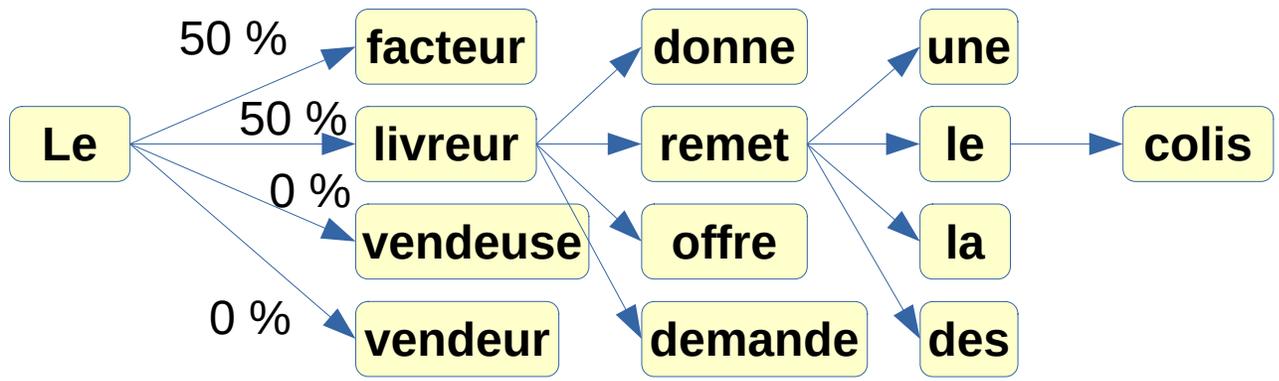
Modélisation générative

ChatGPT et al. sont des modèles génératifs

- On veut, à partir d'un contexte générer un « token »
 - Étant donné une phrase partielle, quel est le prochain mot ?
 - Étant donné des pixels de différentes couleurs, quelle est la couleur du pixel à coté ?
- On apprend des **probabilités** étant donné le contexte

Prompt pour Gemini
Image pour Midjourney

La camionnette est arrivée à 9 heures du matin.



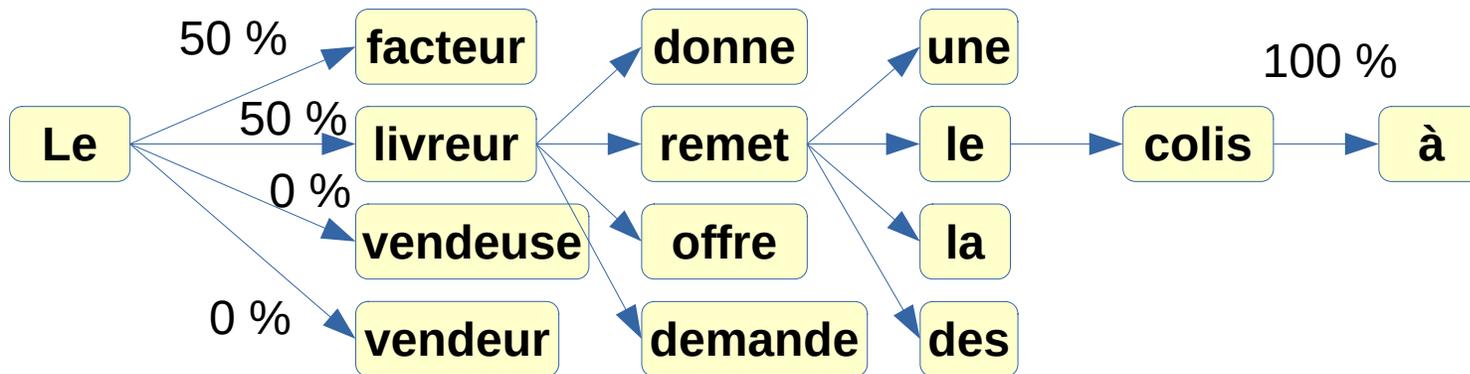
Modélisation générative

ChatGPT et al. sont des modèles génératifs

- On veut, à partir d'un contexte générer un « token »
 - Étant donné une phrase partielle, quel est le prochain mot ?
 - Étant donné des pixels de différentes couleurs, quelle est la couleur du pixel à côté ?
- On apprend des **probabilités** étant donné le contexte

Prompt pour Gemini
Image pour Midjourney

La camionnette est arrivée à 9 heures du matin.



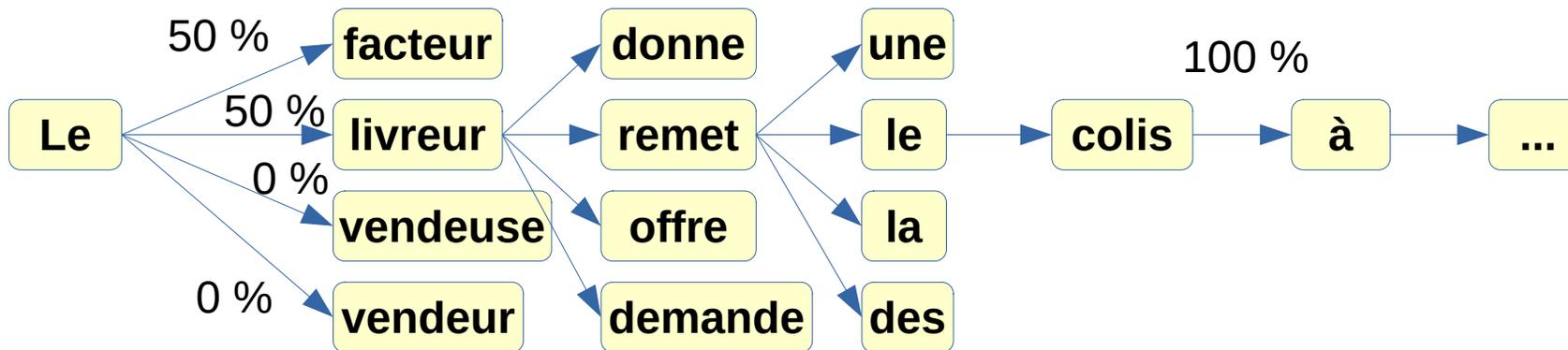
Modélisation générative

ChatGPT et al. sont des modèles génératifs

- On veut, à partir d'un contexte générer un « token »
 - Étant donné une phrase partielle, quel est le prochain mot ?
 - Étant donné des pixels de différentes couleurs, quelle est la couleur du pixel à coté ?
- On apprend des **probabilités** étant donné le contexte

Prompt pour Gemini
Image pour Midjourney

La camionnette est arrivée à 9 heures du matin.



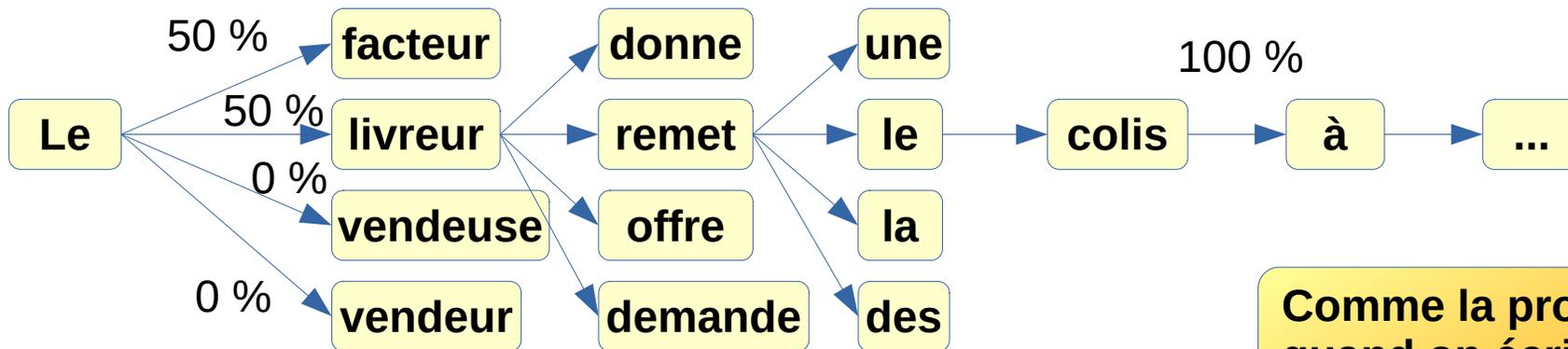
Modélisation générative

ChatGPT et al. sont des modèles génératifs

- On veut, à partir d'un contexte générer un « token »
 - Étant donné une phrase partielle, quel est le prochain mot ?
 - Étant donné des pixels de différentes couleurs, quelle est la couleur du pixel à coté ?
- On apprend des **probabilités** étant donné le contexte

Prompt pour Gemini
Image pour Midjourney

La camionnette est arrivée à 9 heures du matin.



Comme la proposition des mots quand on écrit un message

Les pièges – mauvaises étiquettes

- Système de sélection de receveurs de dons de foie : plus mauvaise chance pour les jeunes
 - Mauvaise étiquette : chance de survie pendant les cinq prochains ans
 - Meilleure chance pour un.e jeune
- Prédiction de risque de pneumonie → admission à l'unité de soins intensifs : les asthmatiques ne sont pas prédits d'être à risque
 - Parce les asthmatiques sont directement admis à l'USI

Les pièges – mauvaises/pas assez de données

- Premiers systèmes de reconnaissance faciale : incapable de reconnaître des personnes racisées
 - Peu de personnes non-blanches dans le jeu de données
- Prédiction du cancer : 100 % précision pendant les tests, inutilisable dans des situations réelles
 - Jeu d'entraînement : dans chaque radiographie cancéreuse, il y avait une règle pour indiquer la taille du tumeur
- Prévision policière/prédiction de récidivisme :
 - Les personnes racisées sont surpoliciées → sur-représentée dans les données historiques
- Traduction automatique : Ne marche très bien pour les langages sous-représentés
 - Pas assez de documents écrits
 - DeepL liste seulement 36 langages (européens + quelques asiatiques)

Les pièges – mauvaises/pas assez de données

- Premiers systèmes de reconnaissance faciale : incapable de reconnaître des personnes racisées
 - Peu de personnes non-blanches dans le jeu de données
- Prédiction du cancer : 100 % précision pendant les tests, inutilisable dans des situations réelles
 - Jeu d'entraînement : dans chaque radiographie cancéreuse, il y avait une règle pour indiquer la taille du tumeur
- Préviation policière/prédiction de récidivisme
 - Les personnes racisées sont surpoliciées → sur-représentée dans les données historiques
- Traduction automatique : Ne marche très bien pour les langages sous-représentés
 - Pas assez de documents écrits
 - DeepL liste seulement 36 langages (européens + quelques asiatiques)

Traduction DeepL

Traduction Google : « sur-policier »

Les pièges – incapacité de raisonner

- Traduction automatique : « The child throws the toy into the pen. » → « L'enfant jette le jouet dans le stylo. »

Les pièges – incapacité de raisonner

- Traduction automatique : « The child throws the toy into the pen. » → « L'enfant jette le jouet dans le stylo. »
 - Pas capable de raisonnement physique

Les pièges – incapacité de raisonner

- Traduction automatique : « The child throws the toy into the pen. » → « L'enfant jette le jouet dans le stylo. »
 - Pas capable de raisonnement physique



Les pièges – incapacité de raisonner

- Traduction automatique : « The child throws the toy into the pen. » → « L'enfant jette le jouet dans le stylo. »
 - Pas capable de raisonnement physique

**Traduction alternative de DeepL :
« L'enfant jette le jouet dans l'enclos. »**



Les pièges – incapacité de raisonner

- Traduction automatique : « The child throws the toy into the pen. » → « L'enfant jette le jouet dans le stylo. »
 - Pas capable de raisonnement physique

Traduction alternative de DeepL :
« L'enfant jette le jouet dans l'enclos. »

- « The child throws the toy into the playpen. » → « L'enfant jette le jouet dans le parc. »



Les pièges – incapacité de raisonner

- Traduction automatique du jouet dans le style

- Pas capable de raisonner

Traduction automatique
« L'enfant jette le jouet »

- « The child throws the toy »

- « L'enfant jette le jouet »

◆ AI Overview 

Based on the size of an Olympic pool and an average African elephant, only one elephant would fit in an Olympic pool with very little room to move around; an Olympic pool is roughly 50 meters long, 25 meters wide, and 2 meters deep, while an average elephant can be around 8.2 to 13 feet (2.5 to 4 meters) tall at the shoulder and weigh between 5,000 to 14,000 pounds. [🔗](#)

Key points to consider:

Elephant size:
An elephant's large body size, particularly its width, would take up most of the pool space. [🔗](#)

Pool dimensions:
An Olympic pool is designed for swimming, not accommodating large animals, so its size limits the number of elephants that could fit. [🔗](#)



Les risques de l'« IA »

- Existe-t-il un risque qu'une IA devienne super-intelligente et mette en danger l'humanité dans un futur proche ?
 - 1) No !
 - 2) C'est le retour de la vieille fable de la création se retournant contre son créateur.
- Mais il y a des risques concret et actuel :
 - Des systèmes qui ne fonctionnent pas correctement et qui affectent des populations vulnérables/marginalisées
 - Des chat bots qui donnent des informations incorrectes

Les risques de l'« IA »

Quand Eliezer Yudkowsky demande des frappes aériennes et nucléaires sur des centres de calculs, c'est très dangereux

- Existe-t-il un risque qu'une IA devienne super-intelligente et mette en danger l'humanité dans un futur proche ?
 - 1) No !
 - 2) C'est le retour de la vieille fable de la création se retournant contre son créateur.
- Mais il y a des risques concret et actuel :
 - Des systèmes qui ne fonctionnent pas correctement et qui affectent des populations vulnérables/marginalisées
 - Des chat bots qui donnent des informations incorrectes

Les risques de l'« IA »

Quand Eliezer Yudkowsky demande des frappes aériennes et nucléaires sur des centres de calculs, c'est très dangereux

- Existe-t-il un risque qu'une IA devienne super-intelligente et mette en danger l'humanité dans un futur proche ?
 - 1) No !
 - 2) C'est le retour de la vieille fable de la création se retournant contre son créateur.
- Mais il y a des risques concret et actuel :
 - Des systèmes qui ne fonctionnent pas correctement et qui affectent des populations vulnérables/marginalisées **Scandale des allocations familiales aux Pays-Bas**
 - Des chat bots qui donnent des informations incorrectes

Les risques de l'« IA »

Quand Eliezer Yudkowsky demande des frappes aériennes et nucléaires sur des centres de calculs, c'est très dangereux

- Existe-t-il un risque qu'une IA devienne super-intelligente et mette en danger l'humanité dans un futur proche ?
 - 1) No !
 - 2) C'est le retour de la vieille fable de la création se retournant contre son créateur.
- Mais il y a des risques concret et actuel :
 - Des systèmes qui ne fonctionnent pas correctement et qui affectent des populations vulnérables/marginalisées
 - Des chat bots qui donnent des informations incorrectes

Scandale des allocations familiales aux Pays-Bas

- Air Canada : Réduction en cas de deuil
- Un avocat qui citaient des cas non-existants
- Accusation inventée d'une agression sexuelle

Les risques de l'« IA » - la consommation des ressources

- Électricité : croissance des émissions de carbone (48 % par rapport à 2019 pour Google, 29 % par rapport à 2020 pour Microsoft)
 - Centres de calcul en Virginia consommeront autant d'électricité que 6 millions de ménages en 2030
 - Une requête à ChatGPT ~ allumer une ampoule pendant 20 minutes
 - Générer une image ~ charger un téléphone à 50 %
- Eau : utilisée pour refroidir les serveurs dans les centres de calcul (2021 → 2022 : 20 % plus [Google], 34 % plus [Microsoft])
 - Entraîner ChatGPT ~ 700.000 litres
 - Une conversation avec ChatGPT (20-50 requêtes) ~ 500 ml

10 fois plus qu'une requête à Google Search

1000 fois plus qu'une requête à Google Search

Les risques de l'« IA » - la consommation des ressources

(2)

- « Ressources humains » invisibles
 - Travailleurs kényans qui étiquetaient des données toxiques (abus sexuels sur les enfants, torture, meurtre, inceste,...) pour < US\$ 2/heure
 - Vénézuéliens, philippins qui étiquetaient des données d'entraînement pour ~ US\$ 1/90 minutes
 - Cruise utilisait 1,5 surveillant / voiture « autonome »
- Données
 - Exploitation des propriétés intellectuelles – les grands acteurs peuvent porter plainte, les artistes indépendants ont du mal de se défendre
 - Exploitation des données sensibles – une modèle entraîné sur des données de santé, de production, de réunion peut les régurgiter
- Argent
 - OpenAI a perdu US\$ 5 Mrd en 2024

Les risques de l'« IA » - la consommation des ressources

(2)

- « Ressources humains » invisibles
 - Travailleurs kényans qui étiquetaient des données toxiques (abus sexuels sur les enfants, torture, meurtre, inceste,...) pour < US\$ 2/heure **Pas de soutien pour leur santé mentale**
 - Vénézuéliens, philippins qui étiquetaient des données d'entraînement pour ~ US\$ 1/90 minutes
 - Cruise utilisait 1,5 surveillant / voiture « autonome »
- Données
 - Exploitation des propriétés intellectuelles – les grands acteurs peuvent porter plainte, les artistes indépendants ont du mal de se défendre
 - Exploitation des données sensibles – une modèle entraîné sur des données de santé, de production, de réunion peut les régurgiter
- Argent
 - OpenAI a perdu US\$ 5 Mrd en 2024

Les risques de l'« IA » - la consommation des ressources

(2)

- « Ressources humains » invisibles

- Travailleurs kényans qui étiquetaient des données toxiques (abus sexuels sur les enfants, torture, meurtre, inceste,...) pour < US\$ 2/heure

Pas de soutien pour leur santé mentale

- Vénézuéliens, philippins qui étiquetaient des données d'entraînement pour ~ US\$ 1/90 minutes

Répétitif, stressant, précaire

- Cruise utilisait 1,5 surveillant / voiture « autonome »

- Données

- Exploitation des propriétés intellectuelles – les grands acteurs peuvent porter plainte, les artistes indépendants ont du mal de se défendre
- Exploitation des données sensibles – une modèle entraîné sur des données de santé, de production, de réunion peut les régurgiter

- Argent

- OpenAI a perdu US\$ 5 Mrd en 2024

Les risques de l'« IA » - la consommation des ressources

(2)

- « Ressources humains » invisibles

- Travailleurs kényans qui étiquetaient des données toxiques (abus sexuels sur les enfants, torture, meurtre, inceste,...) pour < US\$ 2/heure

Pas de soutien pour leur santé mentale

- Vénézuéliens, philippins qui étiquetaient des données d'entraînement pour ~ US\$ 1/90 minutes

Répétitif, stressant, précaire

- Cruise utilisait 1,5 surveillant / voiture « autonome »

Retour du Turc mécanique

- Données

- Exploitation des propriétés intellectuelles – les grands acteurs peuvent porter plainte, les artistes indépendants ont du mal de se défendre
- Exploitation des données sensibles – une modèle entraîné sur des données de santé, de production, de réunion peut les régurgiter

- Argent

- OpenAI a perdu US\$ 5 Mrd en 2024

Les risques de l'« IA » - utilisation

- Les artistes conceptuels, traducteurs techniques, rédacteurs publicitaires, acteurs vocaux, lecteurs pour livres audio, seront remplacés par les modèles génératifs

- Ou forcés d'accepter de pires conditions de travail

- N.B. ni les radiologues (Hinton, 2016) ni les chauffeurs routiers (2017) ont été remplacés

La qualité est moins bonne, mais les modèles sont moins chers/plus rapides.

- Retour de l'effet Eliza : développement des relations « romantiques » avec des chat bots

Qui ne sont pas stables, peuvent être supprimés

- Résultats de recherche incorrects, code bogué etc.

Les modèles ne génèrent pas ce qui est vrai, mais ce qui est probable !

- La recherche en IA a bien progressé depuis 1956
 - Et beaucoup des méthodes développées sont utilisées dans notre vie quotidienne
- On est pourtant toujours très loin d'une vraie « intelligence »
 - Et on n'arrivera pas avec les techniques qui trouvent actuellement le plus de financement
- Le risque d'une super-intelligence qui menace l'humanité est très faible
 - Mais la sur-consommation de ressources, les résultats néfastes par hasard ou intention et la dégradation des conditions de travail sont réels

Merci pour votre attention !

D'autres sources d'informations

- Timnit Gebru, Alex Hanna et al., DAIR : <https://www.dair-institute.org/team/>
- Margaret Mitchell : <https://www.m-mitchell.com>
- Emily M. Bender : <https://faculty.washington.edu/ebender/>
- Gary Marcus : <https://garymarcus.substack.com>
- Arvind Narayanan & Sayash Kapoor : <https://www.aisnakeoil.com>
- Dan McQuillan : <https://www.danmcquillan.org/>
- Ruha Benjamin : <https://www.ruhabenjamin.com>
- Paris Marx : <https://parismarx.com>

On the dangers of stochastic parrots:
Can language models be too big?

Résumé

Click to add Text

Click to add Text