



# QU'EST-CE QUE L'INTELLIGENCE ET COMMENT PEUT-ELLE ÊTRE ARTIFICIELLE ?

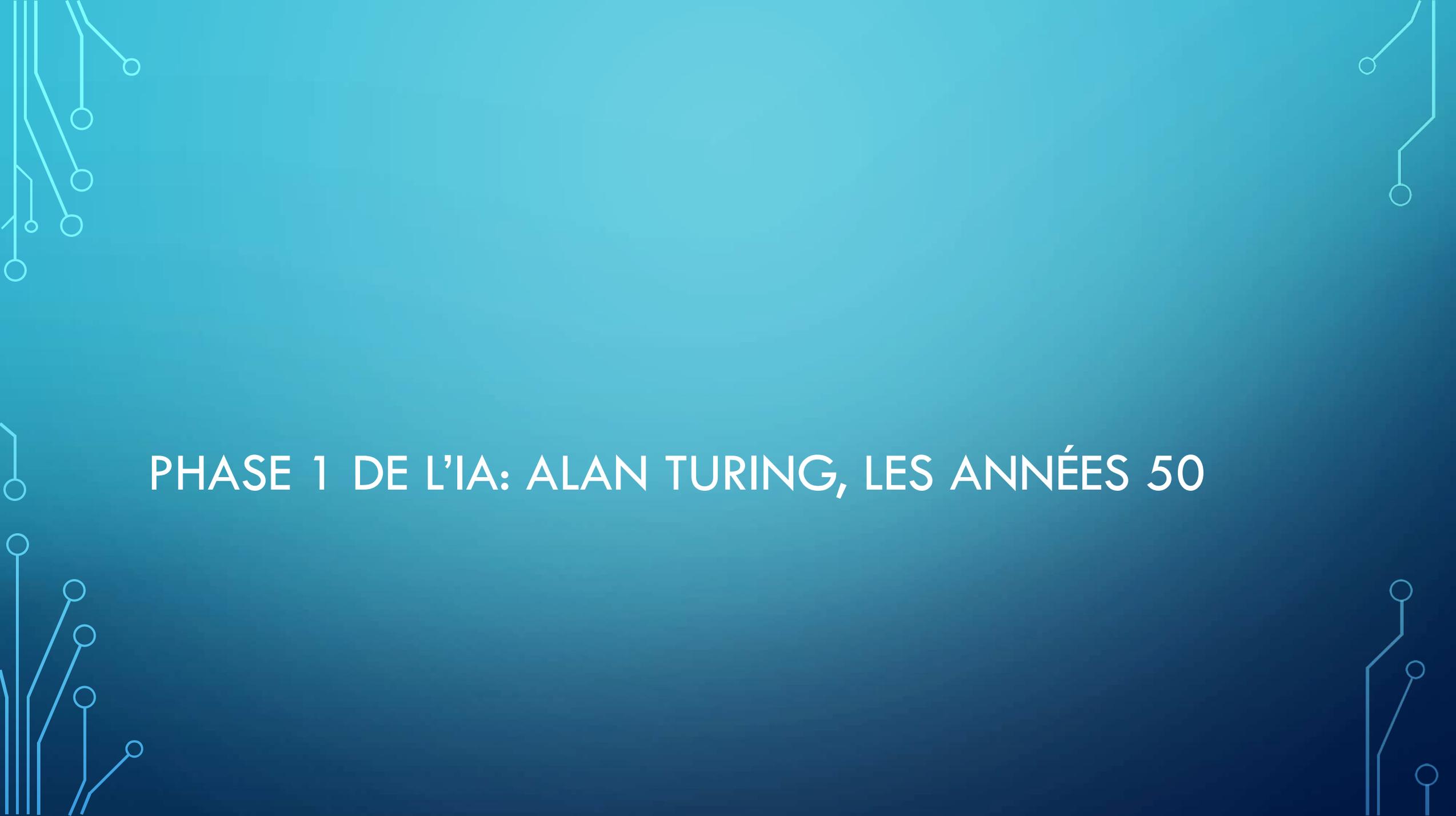
PHI 2320

SÉANCE 12

JONATHAN SIMON

# PROGRAMME

- 1) **PHASE 1 DE L'IA -- Turing**: *La possibilité de machines intelligentes. Comment pouvons-nous tester l'intelligence des machines ?*
- 2) **PHASE 2 DE L'IA -- Symbolisme vs Connectionisme**: *L'intelligence consiste-t-elle à suivre des règles ou à faire des associations et à reconnaître des modèles ?*
- **LA SEMAINE PROCHAINE:**
- 3) **PHASE 3 DE L'IA -- Réseaux neuronaux profonds**: *renouveler le débat entre les empiristes et les innativistes ; comprendre les limites*

The background is a dark blue gradient. In the corners, there are white line-art illustrations of circuit boards or neural networks, with lines connecting to small circles.

# PHASE 1 DE L'IA: ALAN TURING, LES ANNÉES 50

# TURING

- La question: Les systèmes artificiels peuvent-ils penser ??
  - Turing: Oui, Pourquoi pas? (le jeu de l'imitation)
  - Searle: Non -- une syntaxe sans sémantique ne suffit pas (la salle chinois)
  - Dreyfus: Non -- le savoir-faire quotidien n'est pas basé sur la représentation

# ALAN TURING



# MACHINES INFORMATIQUE ET INTELLIGENCE

- 1) Les machines peuvent-elles penser ?
- proposé Turing : il est trop difficile de répondre à cette question. Remplacez-la par :
- 2) Une machine peut-elle être performante (tromper quelqu'un 30 % du temps) au jeu d'imitation ?

# MACHINES INFORMATIQUE ET INTELLIGENCE

- Pourquoi propose-t-il ce remplacement ?
- 1) Il évite un débat philosophique insolubles, le remplace par une question qui reste assez substantielle
- 2) Comportementalisme / Positivisme ? C'était 1950...
- 3) Théorie cartésienne de la primauté de la preuve de la capacité linguistique (rappel de Descartes vs. Romanes et Huxley)

# MACHINES INFORMATIQUE ET INTELLIGENCE

- Le jeu de l'imitation :
- Un examinateur, deux joueurs. Un joueur est humain, l'autre est une machine. Les joueurs sont dans des salles séparées, seuls des messages écrits sont échangés. L'examineur peut demander n'importe quoi à l'un ou l'autre des participants, le but étant de deviner qui est l'humain. L'objet du jeu est de tromper l'examineur en lui faisant croire que l'autre joueur est la machine.

# MACHINES INFORMATIQUE ET INTELLIGENCE

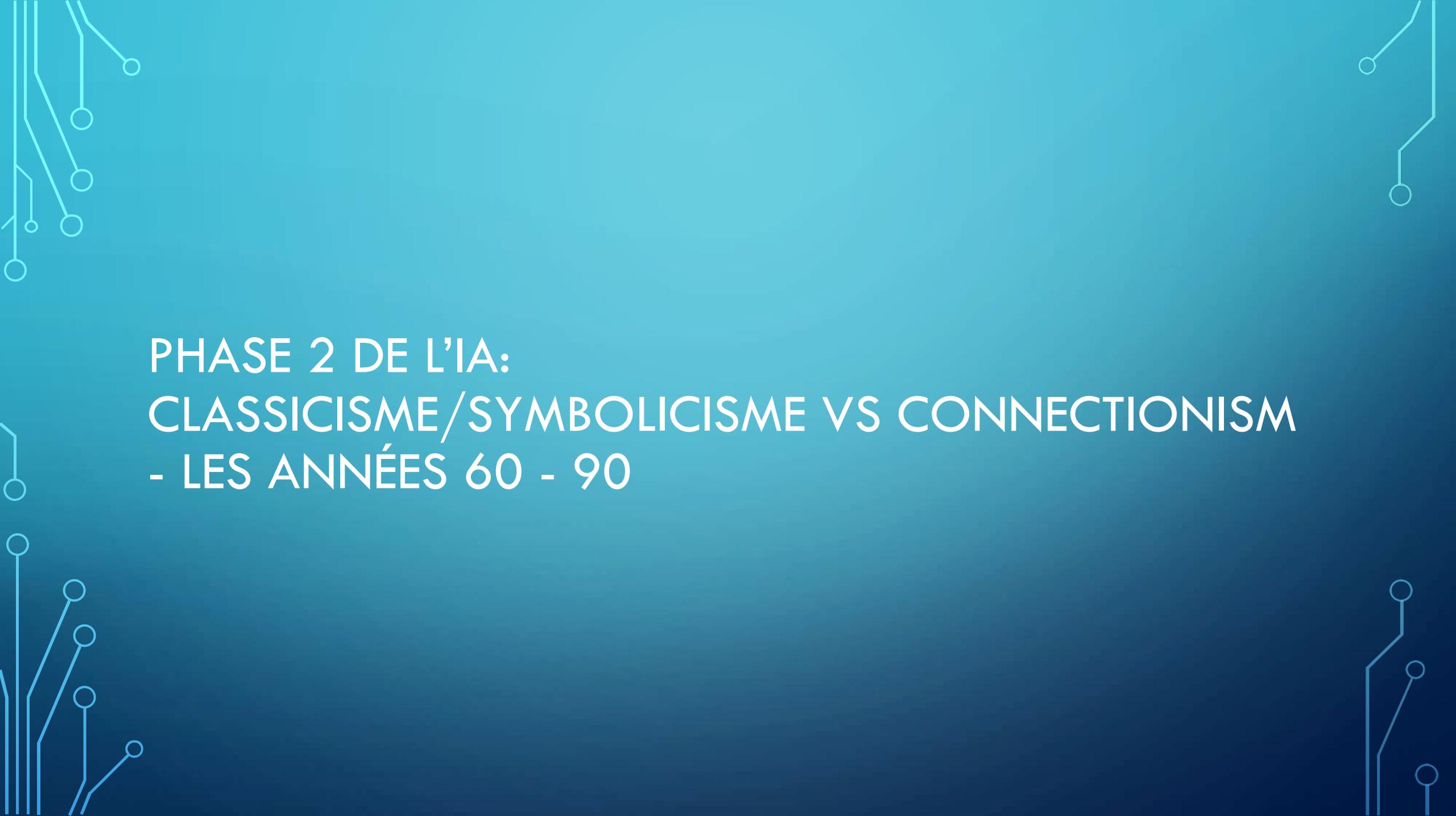
- Turing précise ensuite qu'il pense que les ordinateurs artificiels, construits à partir des machines de Turing, pourraient être capables de passer ce test dans 50 ans (par exemple, vers 2001....)

# MACHINES INFORMATIQUE ET INTELLIGENCE

- Principales objections :
- Searle et la chambre chinoise :
- *L'homme dans la pièce mettrait en œuvre une machine de Turing et passerait le test d'imitation mais ne parlerait pas chinois*

# MACHINES INFORMATIQUE ET INTELLIGENCE

- Principales objections :
- Dreyfus et le suivi des règles :
- ***L'intelligence et l'expertise ne proviennent pas du suivi des règles. Les règles sont des roues d'entraînement dont les experts se débarrassent, et non la base de la rationalité. – savoir faire vs connaissance que  
(il s'inspire ici de Heidegger et du tarif Wittgenstein).***

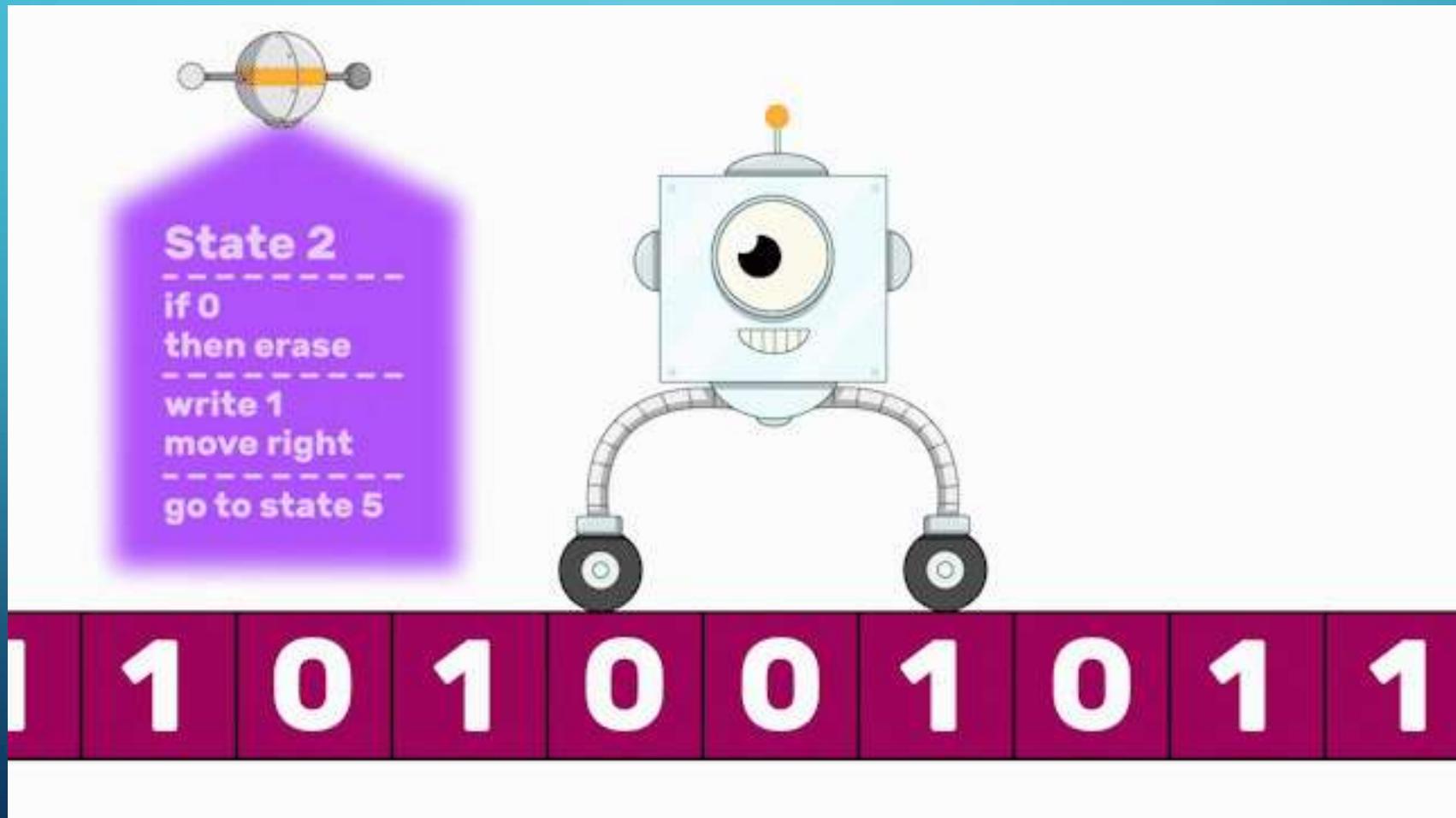
The background is a dark blue gradient. In the corners, there are decorative white line-art elements resembling circuit traces or neural network connections, with small circles at the end of the lines.

PHASE 2 DE L'IA:  
CLASSICISME/SYMBOLICISME VS CONNECTIONISM  
- LES ANNÉES 60 - 90

# L'IA CLASSIQUE

- Systemes de traitement de symboles (GOFAI)
- Les états d'une machine de Turing représentent directement et explicitement les pensées, les plans, les instructions

# MACHINES DE TURING (RAPPEL)



# MACHINES DE TURING (RAPPEL)

TABLE 1

	$S_1$	$S_2$
Input : pièce de 5 cents	N'émettre aucun output Passer en $S_2$	Emettre un coca-cola Passer en $S_1$
Input : pièce de 10 cents	Emettre un coca-cola Rester en $S_1$	Emettre un coca-cola et une pièce de 5 cents Passer en $S_1$

# L'IA CLASSIQUE

- Systemes de traitement de symboles (GOFAI)
- Les états d'une machine de Turing représentent directement les pensées, les plans, les instructions
- L'IA précise beaucoup de faits et de règles formelles : une théorie formelle de toute la réalité du sens commun

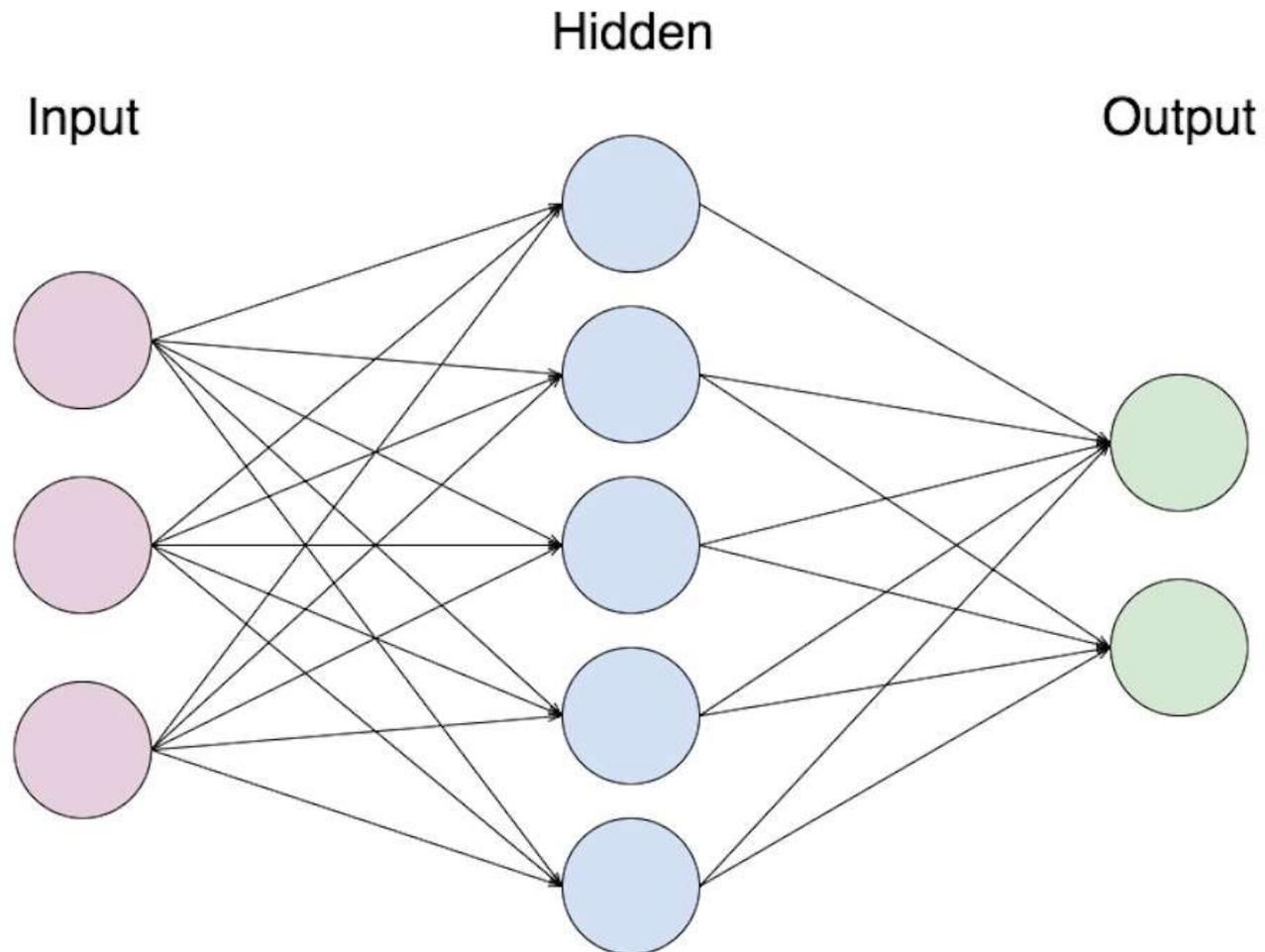
# L'IA CLASSIQUE

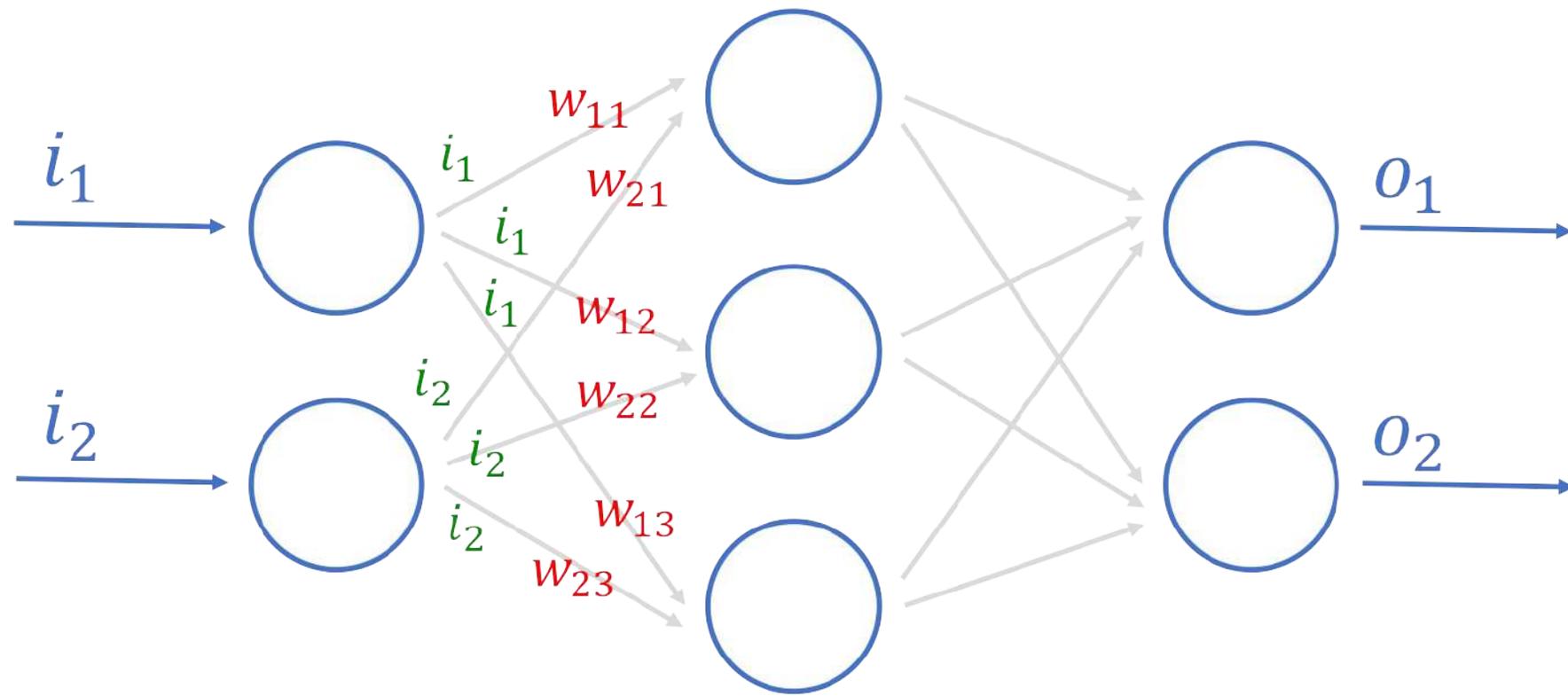
- Systemes de traitement de symboles (GOFAI)
- (Remarque : compris comme un langage machine, ce qui est explicite dans un état n'est toujours qu'une série de 1 et de 0... les règles explicites ou symboliques peuvent être comprises soit comme intégrées dans les règles du tableau machine, soit comme exprimées dans une interprétation de plus haut niveau des 1 et des 0... les symboles composés de 1 et de 0).

# CONNECTIONNISME

- L'approche rivale, née dans les années 60 et 70.
- **Réseaux de neurones artificielles** : structure informatique modelée sur les réseaux de neurones biologiques
- **L'unité de calcul de base est modelée sur le tir des neurones** (beaucoup d'opérations très simples et subsymboliques - pas de représentation explicite des faits ou des règles au niveau de la pensée humaine, au lieu de cela, ceux-ci sont codés implicitement dans des poids neuronaux)

# CONNECTIONNISME



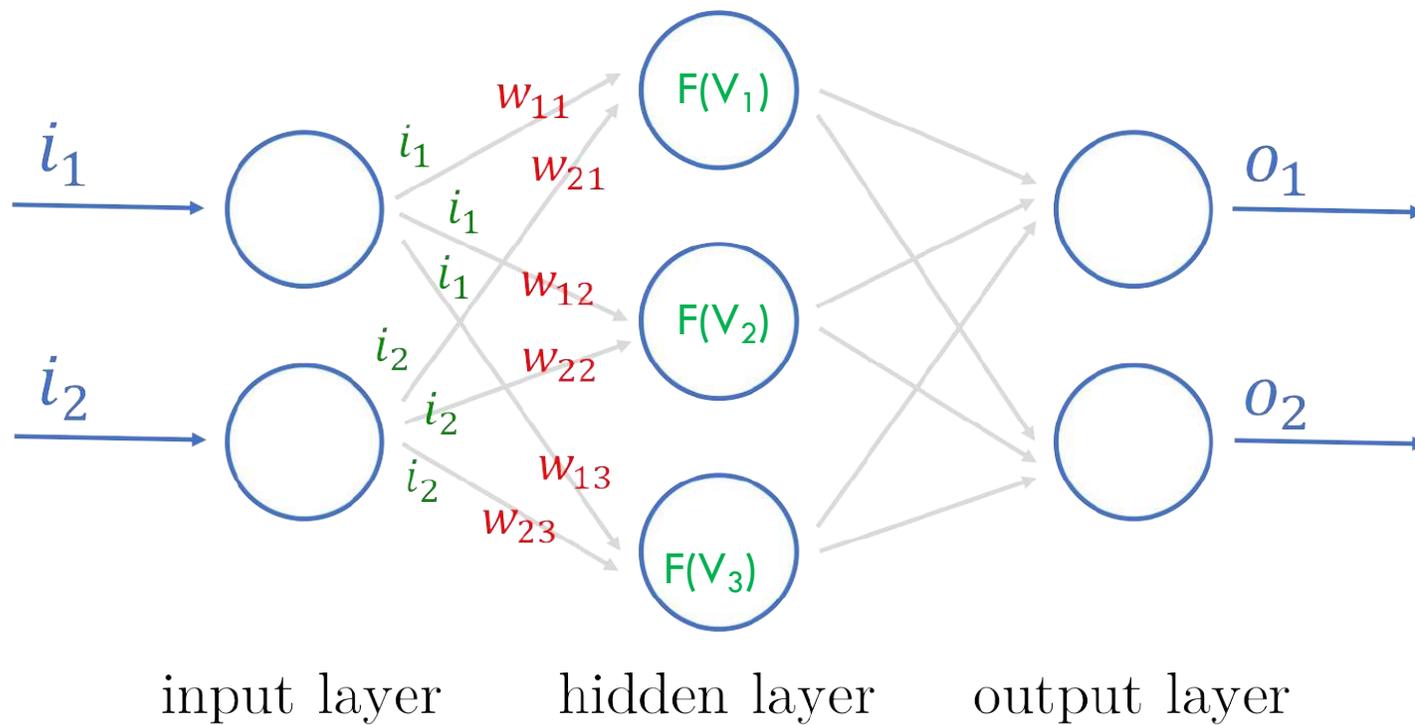


input layer

hidden layer

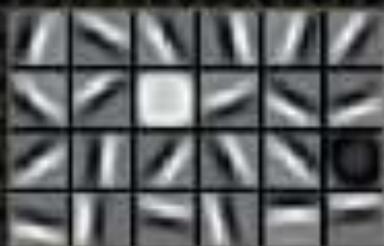
output layer

$$\begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ w_{13} & w_{23} \end{bmatrix} \cdot \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} (w_{11} \times i_1) + (w_{21} \times i_2) \\ (w_{12} \times i_1) + (w_{22} \times i_2) \\ (w_{13} \times i_1) + (w_{23} \times i_2) \end{bmatrix}$$

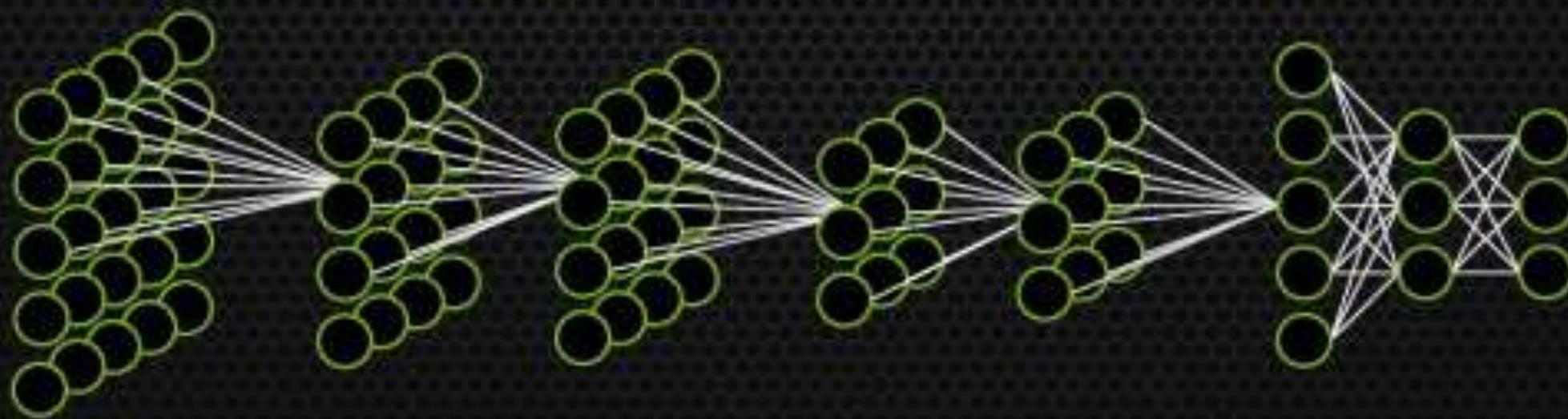


# CONNECTIONISME

- La vision :



Image

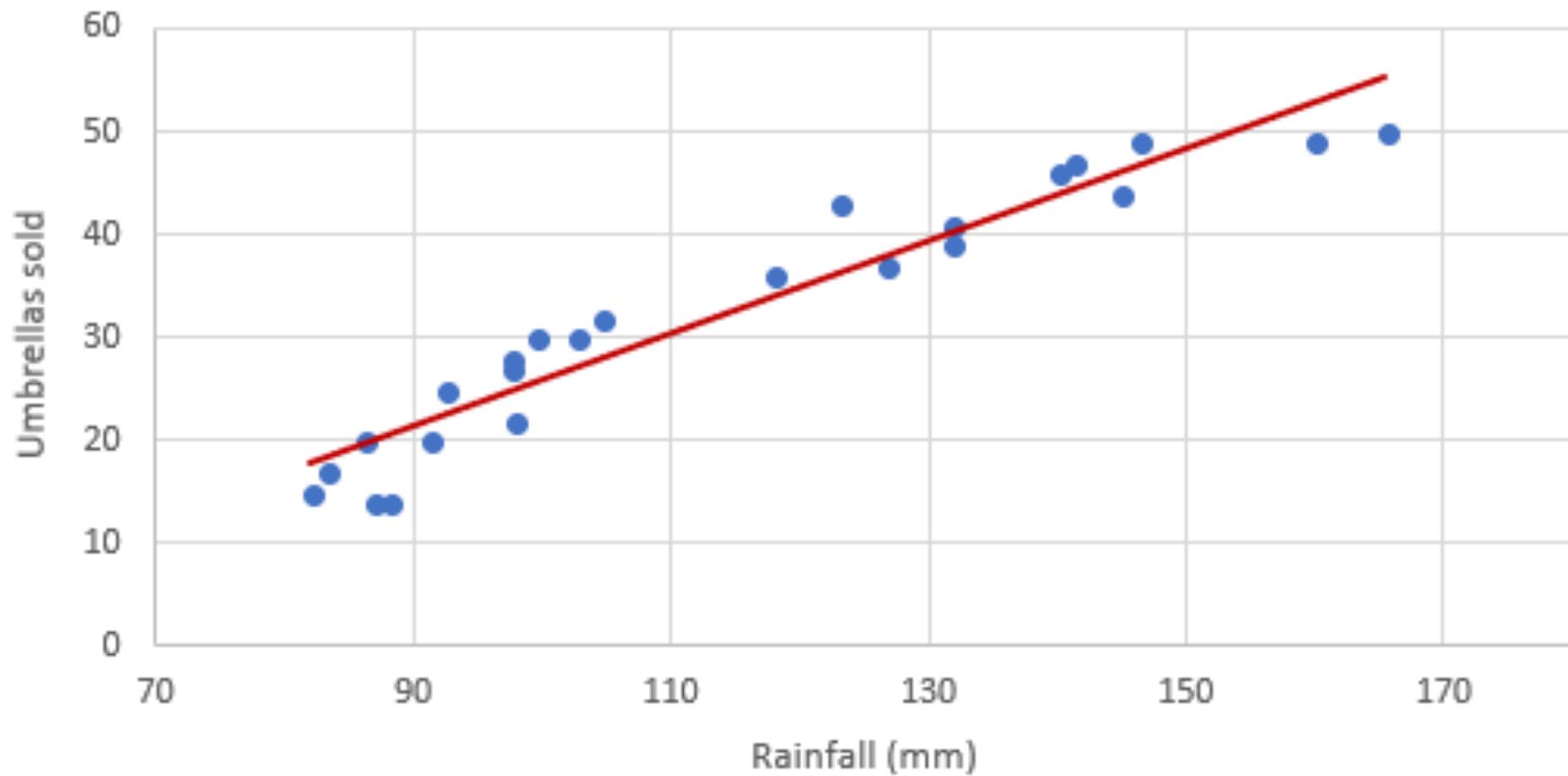


"Sara"

# CONNECTIONNISME

- Ces méthodes sont toutes, en fait, des moyens d'approximer des fonctions (distributions de données). La régression linéaire est la plus simple :

## Linear regression

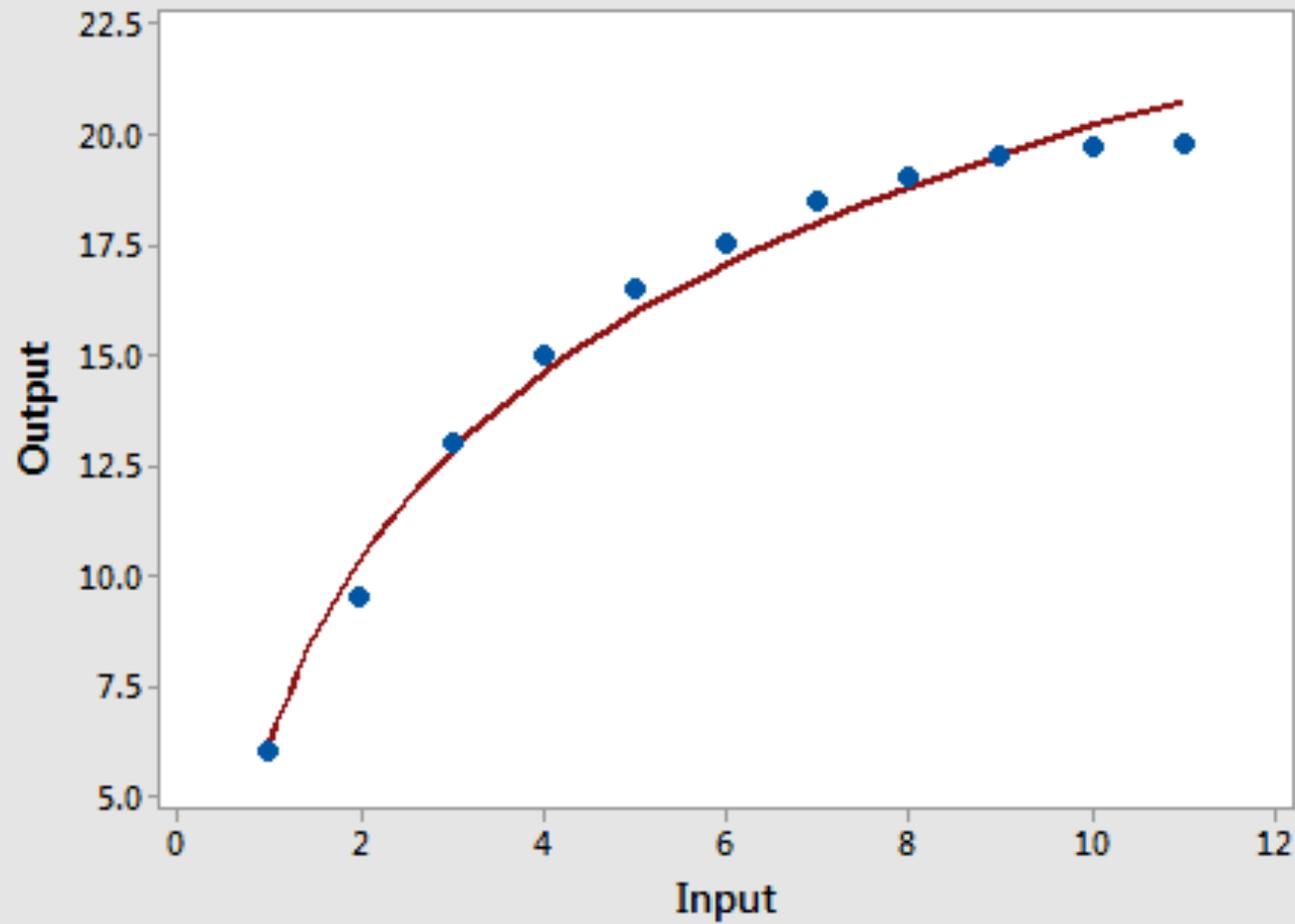


# CONNECTIONNISME

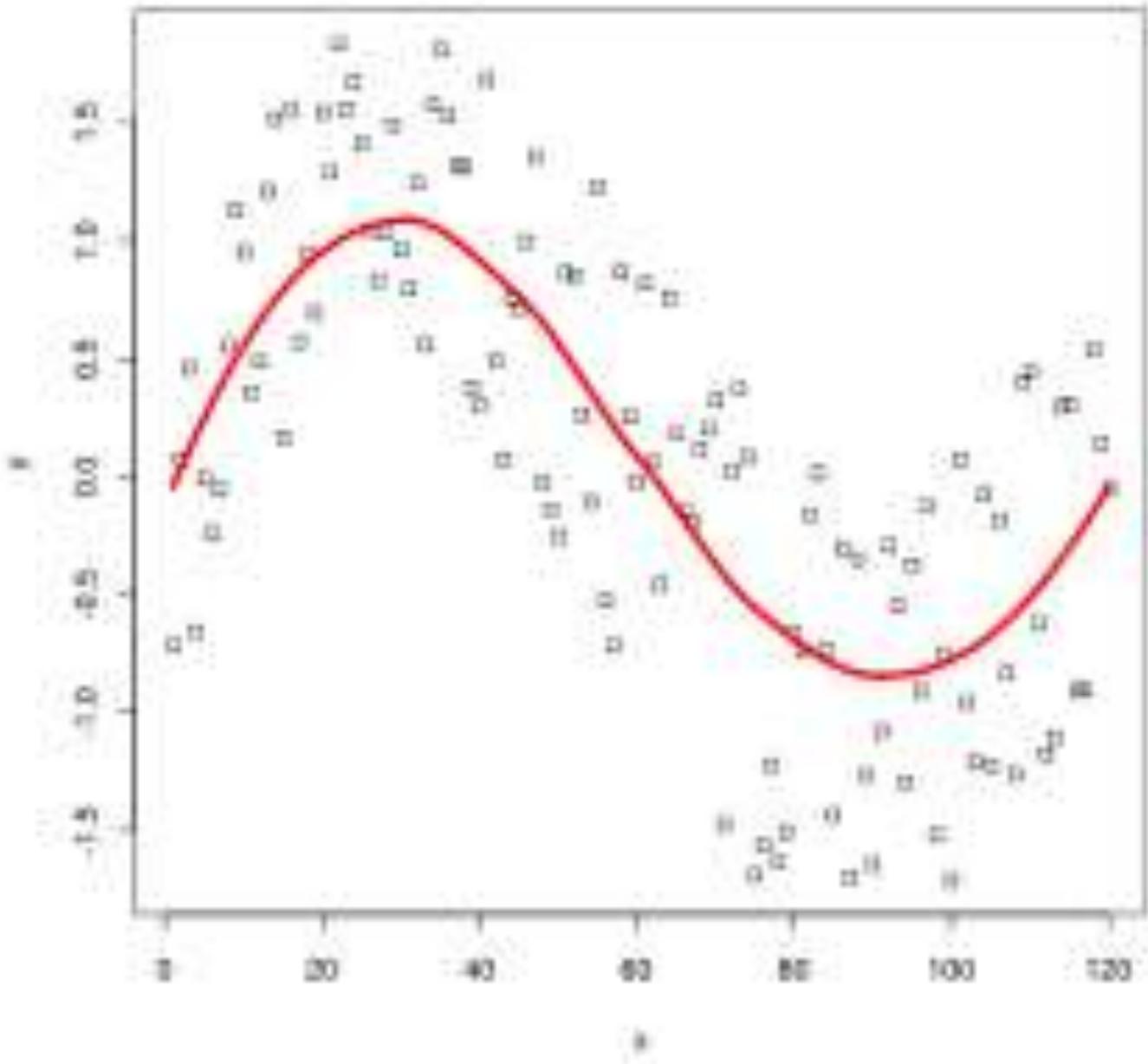
- Plus on ajoute de termes, plus l'approximation peut devenir sophistiquée :

### Fitted Line Plot

$$\text{Output} = 6.099 + 14.06 \log_{10}(\text{Input})$$



S	0.565293
R-Sq	98.6%
R-Sq(adj)	98.5%



# Linear Regression Curve

Daily Chart - E-mini S&P 500 Future (ES)



Commodity.com - all rights reserved

Copyright © 2008

# CONNECTIONNISME

- Il existe un théorème, le théorème d'approximation universelle, qui dit qu'un réseau neuronal représenté comme ci-dessus peut approximer n'importe quelle fonction : plus on veut que l'approximation soit compliquée, plus on ajoute de neurones...

# CONNECTIONNISME

- Note que tu peux utiliser une telle approximation pour faire des prédictions, c'est-à-dire pour remplir les cas pour lesquels tu n'as pas encore de données

# CONNECTIONNISME

- Notez: il s'agit d'un débat sur le logiciel (*software*), pas sur le matériel (*hardware*):
- le même ordinateur peut mettre en œuvre les deux (et en fait, ton ordinateur le fait !)... cela peut toujours être une machine de Turing : la question est de savoir comment comprendre ses états
- - les «neurones» des réseaux neuronaux ne sont que des représentations numériques, *modélées* sur des neurones.

# CONNECTIONNISME

- IA Classique: Les règles du tableau et les états de la machine de bas niveau (au niveau des 1 et des 0), lorsqu'ils sont lus à un niveau supérieur, correspondent à des mots ou à des déclarations et à des règles pour raisonner avec eux.
- Connectionisme: les états internes sont des collections de chiffres (matrices, tenseurs) qui représentent des groupes de neurones (une matrice représente une couche, une ligne de la matrice représente un neurone), et les opérations (généralement la multiplication des matrices) représentent des événements de tir neuronal : des événements où chaque neurone transmet un signal unique aux neurones auxquels il est directement connecté dans la couche suivante

# CONNECTIONNISME

- Le point important est que les états de machines internes pertinents pour l'IA classique sont des états qui codent directement, peut-être explicitement des faits et des principes sur le monde, écrits sous forme symbolique, et les opérations de calcul formelles sont censées refléter le raisonnement (ainsi, le raisonnement est une question de manipulation purement syntaxique des symboles)
- Le connexionnisme le nie et considère que les représentations explicites sont inférieures au niveau des concepts humains, des règles, etc. : elles émergent de l'activité complexe des neurones (les représentations plus simples et explicites).

# CONNECTIONNISME

- On peut considérer que la distinction ultime se situe entre la déduction (logique) et l'inférence statistique.
- L'IA classique fonctionne par déduction logique : sa «cognition» correspond à des déductions données par les règles (sa table machine) et sa «pensée» ou «intention» (son état machine).
- Le connexionnisme fonctionne par inférence statistique : il trouve la «courbe» la plus simple pour s'adapter aux données, puis classe les nouvelles choses en fonction de leur position sur cette courbe (nous y reviendrons dans les semaines à venir).

# COMPARER LES DEUX:

- Trier les chats des chiens :
- Supposons que nous ayons des images en pixels que nous voulons alimenter à un système : soit une image de chat, soit une image de chien
- GOFAI: Il faut d'abord des règles pour convertir les motifs de pixels en bords, puis les motifs de bords en formes, puis les motifs de formes en, par exemple, oreille de chien, nez de chat, etc. A la dernière étape, vous avez terminé. Mais essayez de penser à une règle que vous pourriez écrire explicitement, pour déduire quand une forme est en forme d'oreille de chien plutôt que de chat...

# COMPARER LES DEUX:

- Trier les chats des chiens :
- Supposons que nous ayons des images en pixels que nous voulons alimenter à un système : soit une image de chat, soit une image de chien
- Réseaux neuronaux : vous l'entraînez sur beaucoup, beaucoup d'images (pour lesquelles vous lui dites si c'est un chien ou un chat). Ensuite, il ajuste ses poids (les valeurs exactes dans les matrices), en ajustant une frontière très complexe (côté chien, côté chat) dans un espace de représentation de valeurs de pixels. Ensuite, lorsque vous lui montrez une nouvelle image, il repère simplement de quel côté de la limite il se trouve.

# COMPARER LES DEUX:

- Trier les chats des chiens :
- Supposons que nous ayons des images en pixels que nous voulons alimenter à un système : soit une image de chat, soit une image de chien
- Réseaux neuronaux : vous l'entraînez sur beaucoup, beaucoup d'images (pour lesquelles vous lui dites si c'est un chien ou un chat). Ensuite, il ajuste ses poids (les valeurs exactes dans les matrices), en ajustant une frontière très complexe (côté chien, côté chat) dans un espace de représentation de valeurs de pixels. Ensuite, lorsque vous lui montrez une nouvelle image, il repère simplement de quel côté de la limite il se trouve.

## COMPARER LES DEUX:

- Règles contre associations (Rationalisme vs Empirisme)
- IA symbolique (classique) : règles. Rationalisme:
- nous sommes câblés pour faire du raisonnement (logique) basé sur des règles, qui utilise souvent la structure syntaxique des représentations : nous déduisons P de «P et Q» non pas parce que nous avons remarqué (dans notre expérience) que le premier est généralement vrai lorsque le second l'est, mais parce que nous comprenons que cela fait partie de ce que «et» signifie

## COMPARER LES DEUX:

- Règles contre associations (Rationalisme vs Empirisme)
- IA connexioniste : associations. Empirisme:
- Notre compréhension est basée sur un apprentissage associatif (empirique) sensible au contexte. Les règles logiques «dures et rapides» sont l'exception (plutôt que la règle). En général, nous raisonnons de manière probabiliste et nous comprenons de manière associative.

## COMPARER LES DEUX:

- Intuitivement, il pourrait être logique de donner une place aux deux : le connexionnisme pour l'intuition, la perception, la cognition subpersonnelle, le raisonnement symbolique pour la pensée consciente.
- (Smolensky: le processeur intuitif contre le processeur conscient)

## COMPARER LES DEUX:

- Mais il est difficile de voir comment intégrer les deux, (ou du moins c'était le cas pendant la phase 2).
- ... des problèmes pour l'approche symbolique dans l'explication du niveau «processeur intuitif» (cognition de niveau 1, dans le langage de la théorie du double processus)
- ... des problèmes pour l'approche connexionniste lorsqu'il s'agit d'expliquer le niveau du «processeur conscient» (cognition de niveau deux, dans le langage de la théorie du double processus).

## COMPARER LES DEUX:

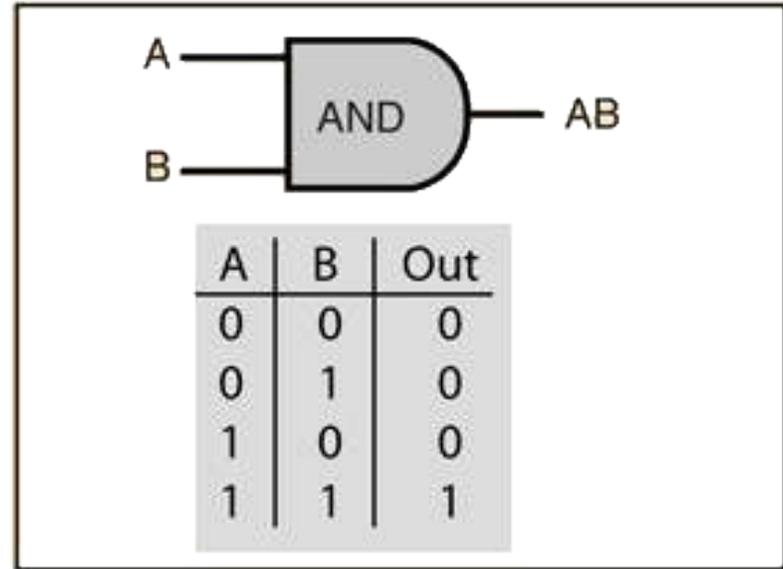
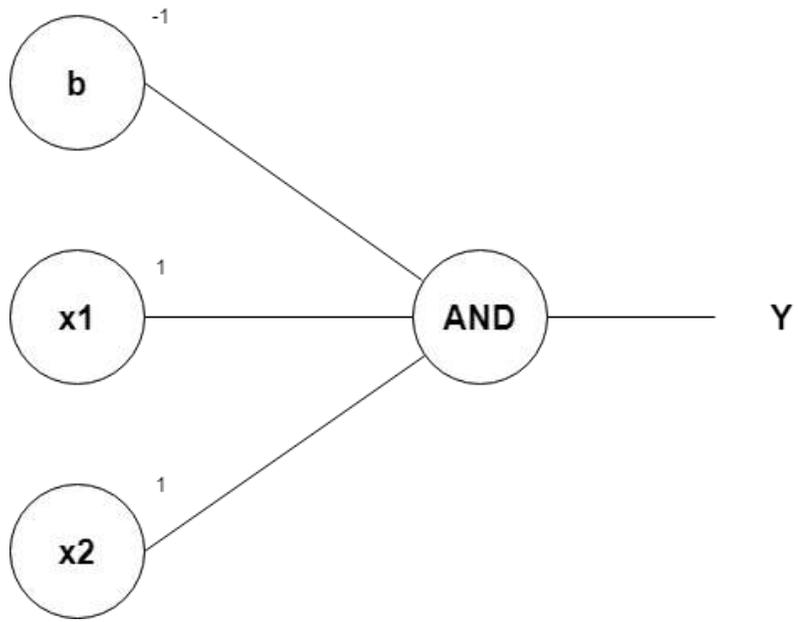
- ... des problèmes pour l'approche symbolique dans l'explication du niveau «processeur intuitif» :
- Il est très difficile d'écrire des règles explicites pour distinguer les photos de chats des photos de chiens ! (Dreyfus, Smolensky...)

## COMPARER LES DEUX:

- ... des problèmes pour l'approche connectioniste dans l'explication du niveau «processeur intuitif» :
- Les systèmes statistiques peuvent se rapprocher des règles déductives mais ne les capturent jamais vraiment (Fodor, Pylyshyn, Gary Marcus)

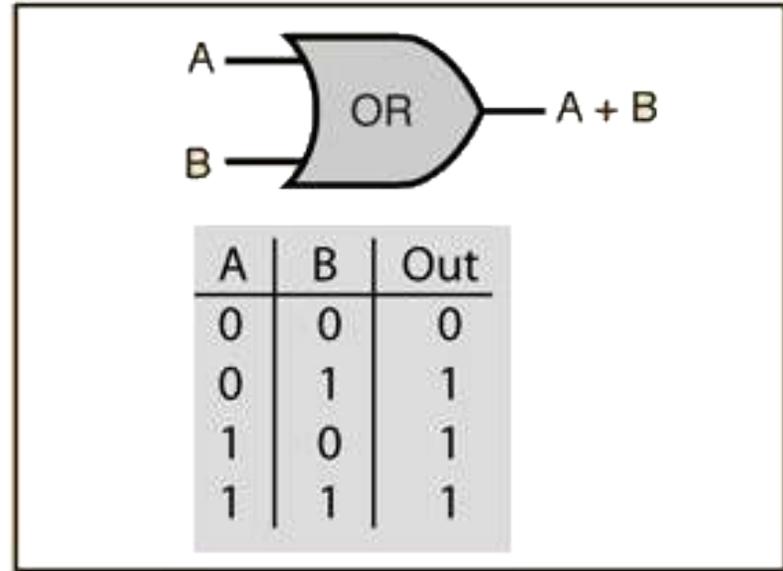
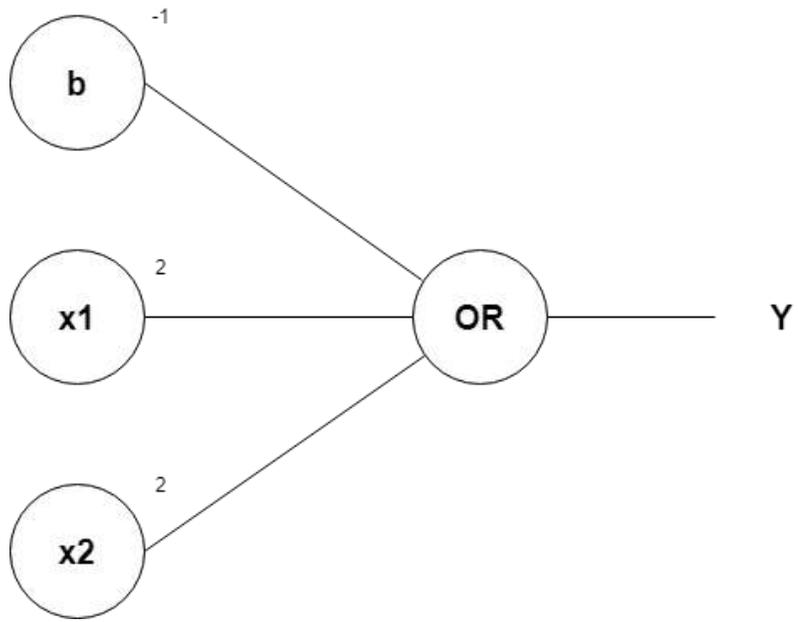
## COMPARER LES DEUX:

- ... des problèmes pour l'approche connectioniste dans l'explication du niveau «processeur intuitif» :
- À noter: Le problème n'est pas que les réseaux neuronaux ne peuvent pas modéliser des règles déductives. On peut mettre en œuvre une machine de Turing avec un réseau neuronal: Voici une «and-gate» et une «or-gate»



# SMOLENSKY

“ET” (AND)



A	B	Out
0	0	0
0	1	1
1	0	1
1	1	1

# SMOLENSKY

“OU” (OR)

## COMPARER LES DEUX:

- ... des problèmes pour l'approche connectionniste dans l'explication du niveau «processeur intuitif» :
- L'objection ne peut donc pas être qu'un modèle de réseau neuronal statistique ne peut pas accueillir de règles du tout...

## FODOR ET PYLYSHYN:

- L'argument de la systématique : un défi plus sophistiqué au connexionnisme.
- Pas simplement : il ne peut pas expliquer la règle déductive de la conjonction (le et-porte).
- Plutôt : un modèle qui est finalement statistique, plutôt que juste une implémentation neuronale d'un processeur symbolique classique, ne sera pas capable d'utiliser ces règles de manière systématique, compositionnelle et productif...

# FODOR ET PYLYSHYN

- L'idée centrale : la pensée (et le langage) sont productifs, systématiques et compositionnels.
- Pour que ces caractéristiques soient réellement mises en œuvre, il doit y avoir une structure de partie et d'ensemble parmi les éléments syntaxiques qui représentent les pensées, et cette structure doit être causalement pertinente et aider à expliquer les processus de pensée
- Mais les systèmes connexionnistes n'ont pas ce genre de structure partie-et-ensemble (au niveau cognitif)

# FODOR ET PYLYSHYN

- Prémisse 1) La pensée (et le langage) sont productifs, systématiques et compositionnels.
- Prémisse 2) Si la pensée est productif, systématique et compositionnel, il a une structure (causalement pertinent) de partie et d'ensemble
- Prémisse 3) Mais les systèmes connexionnistes n'ont pas une structure (causalement pertinent) de partie et d'ensemble
- Conclusion) Donc, la pensée n'est pas connexionniste

# FODOR ET PYLYSHYN

- Prémisse 1) La pensée (et le langage) sont productifs, systématiques et compositionnels.
  - Le débat ici : la pensée a-t-elle vraiment ces caractéristiques, ou ne s'agit-il que d'approximations utiles ?

# FODOR ET PYLYSHYN

- La pensée (et le langage) sont productifs, systématiques et compositionnels:
- Productivité : nous pouvons générer et comprendre des phrases arbitrairement longues
- Systematicité : si nous comprenons une phrase en utilisant les mots A, B et C, nous pouvons comprendre toutes les phrases utilisant ces mots. (Si vous comprenez "Jean aime Marie", vous pouvez comprendre "Marie aime Jean")
- Compositionnalité : les mots apportent la même contribution au sens indépendamment de leur contexte

# FODOR ET PYLYSHYN

- La productivité et la compositionnalité sont controversées. Peut-on vraiment comprendre des phrases de 400 mots ?
- Les mots ont-ils vraiment la même signification dans tous les contextes ? (grain de café, tasse de café, plante de café...)

# FODOR ET PYLYSHYN

- Il est beaucoup plus difficile de nier le systematicité
- si je comprends "Jean aime Marie", je peux comprendre "Marie aime Jean".

# FODOR ET PYLYSHYN

- Prémisse 2) Si la pensée est productif, systématique et compositionnel, il a une structure (causalement pertinent) de partie et d'ensemble
  - Le débat ici : y a-t-il des moyens (connectionnistes) de réaliser ces caractéristiques, sans une structure (causalement pertinent) de partie et d'ensemble?

# FODOR ET PYLYSHYN

- Prémisse 3) Mais les systèmes connexionnistes n'ont pas une structure (causalement pertinent) de partie et d'ensemble
  - Le débat ici : est-il vrai qu'aucun système connexionniste n'a la structure adéquate (sauf s'il s'agit en fait d'une simple mise en œuvre d'une architecture classique) ?

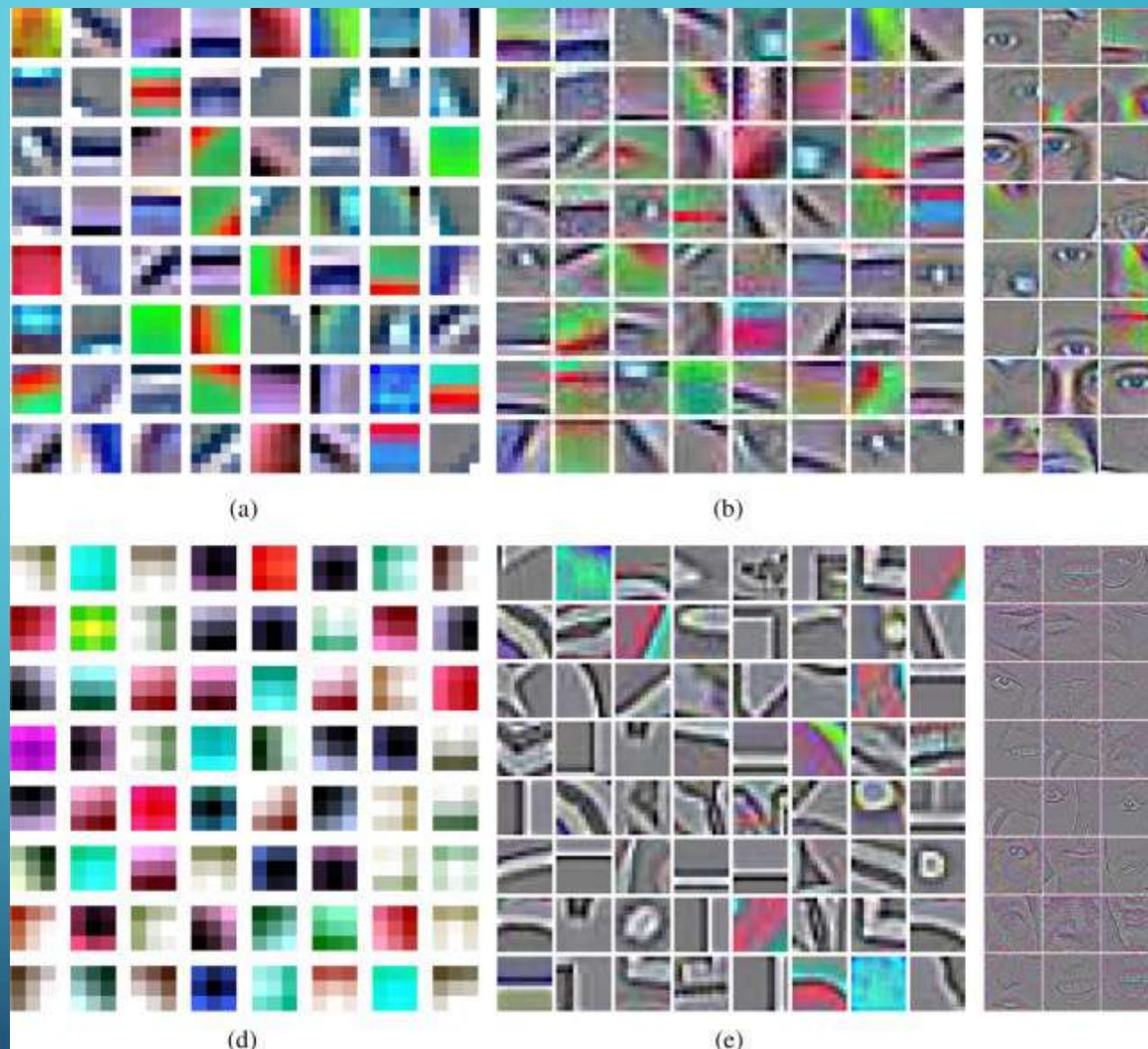
# FODOR ET PYLYSHYN

- Qu'est-ce qu'on veut dire par « une structure (causalement pertinent) de partie et d'ensemble » ?
- Dans le traitement classique, les contenus comme "vache brune" ou "vache et brune" sont représentés par des symboles composés. Ils ont des parties. Le symbole de la "vache" est une partie du symbole de la "vache brune". Les règles de traitement classique peuvent fonctionner sur cette structure, c'est-à-dire que si vous pensez que "P et Q", alors concluez que "P" peut être mis en œuvre de manière syntaxique : si vous acceptez un contenu qui est un composé de deux contenus entourant le symbole "et", alors acceptez le contenu atomique à gauche.

# FODOR ET PYLYSHYN

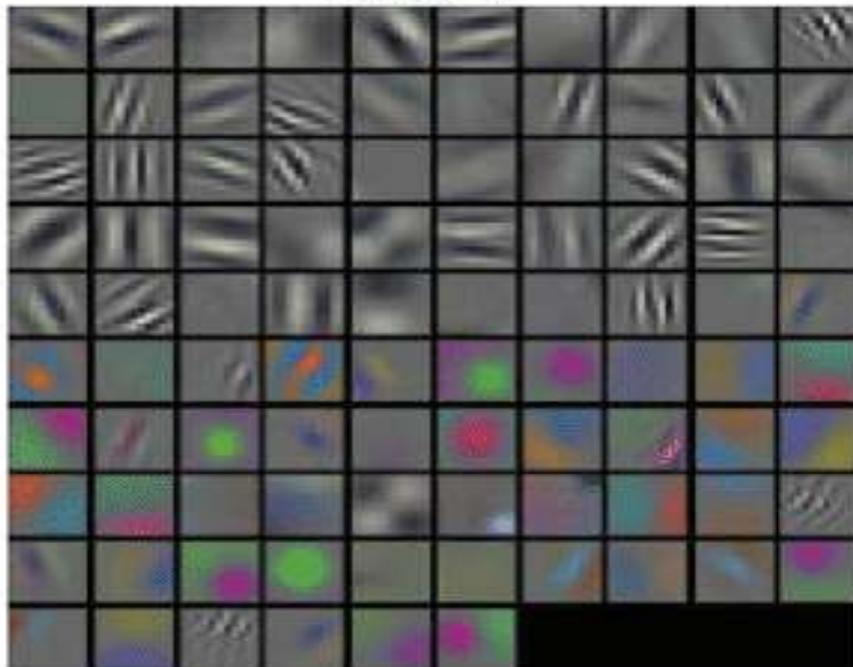
- Qu'est-ce qu'on veut dire par « une structure (causalement pertinent) de partie et d'ensemble » ?
- Dans le traitement connexionniste, Les contenus sont représentés par des modèles de tir de neurones ou de groupes de neurones. L'idée étant que le neurone ou l'amas de neurones tire beaucoup pour signaler la présence de ce qu'il représente.
- En général, ce n'est pas parce que le neurone N1 signale la vache et le N2 le marron que l'on peut raisonner efficacement avec la "vache marron", mais parce qu'il faut un autre neurone N3 pour le signaler. (donc pas de structure de composition automatique). C'est précisément ce que font les Deep Nets : ils ajoutent des couches supplémentaires de neurones pour représenter plus de contenu.

# FODOR ET PYLYSHYN

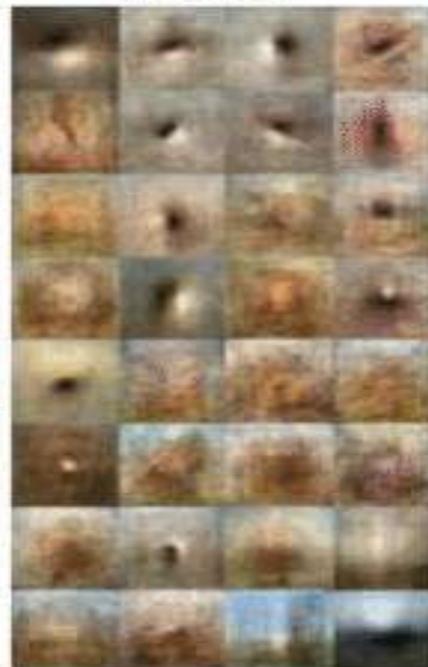


ImageNet-CNN

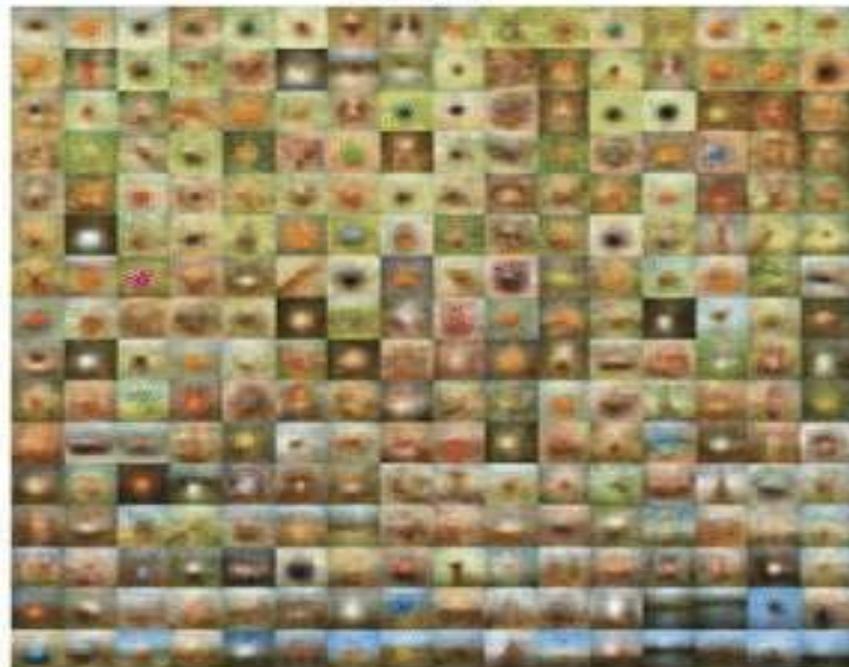
Conv 1



Pool 2



Pool 5

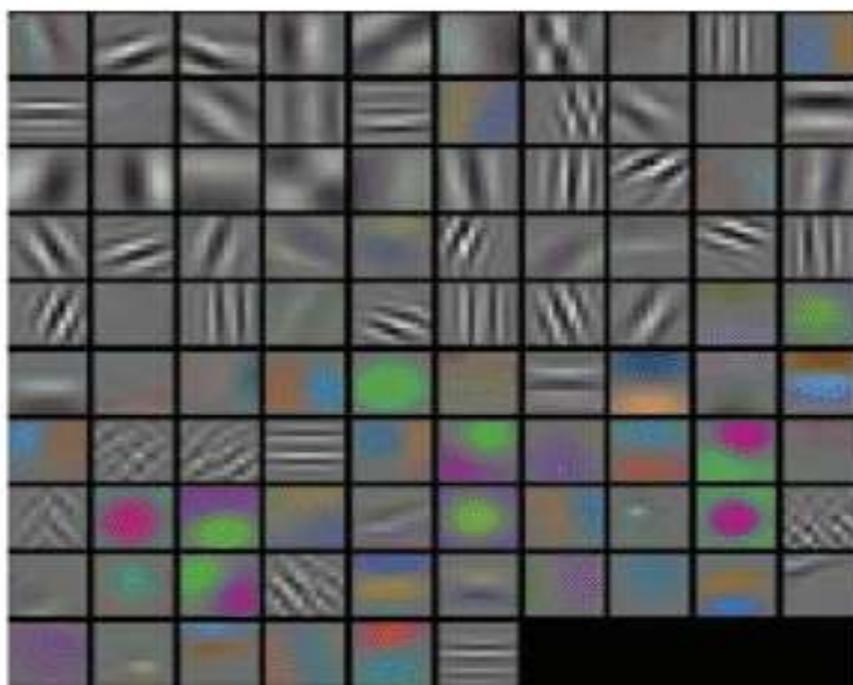


FC 7



Places-CNN

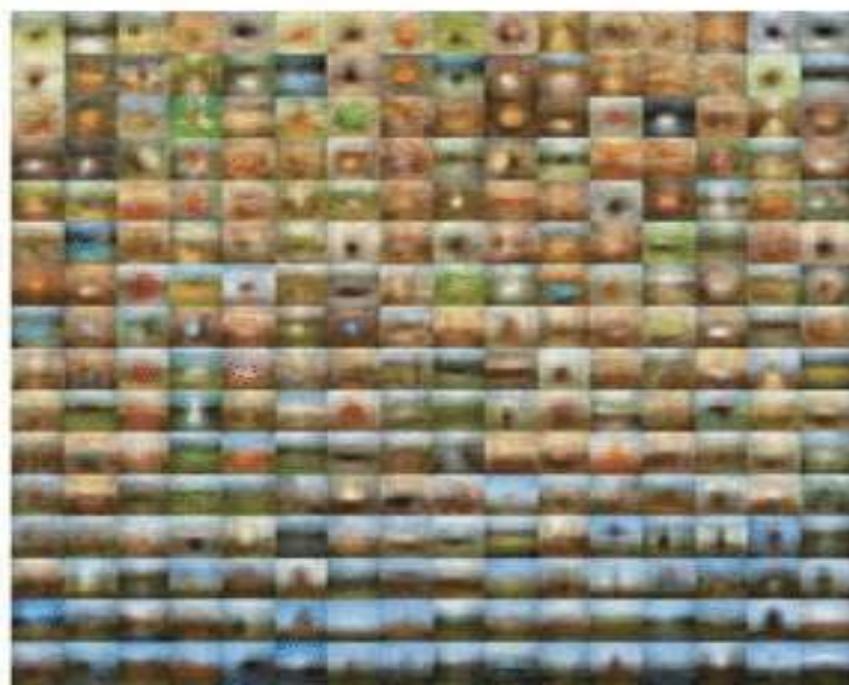
Conv 1



Pool 2



Pool 5



FC 7



# FODOR ET PYLYSHYN

- Il est beaucoup plus difficile de nier le systematicité
- si je comprends "Jean aime Marie", je peux comprendre "Marie aime Jean".
- Mais attention : même si nous admettons que le tir du neurone pour le brun et le tir pour la vache constituent une représentation composite de la vache brune, nous avons ici un problème : considerez: «Jean aime Marie»: comment distinguer entre cela et «Marie aime Jean»)

# FODOR ET PYLYSHYN

- Nous pouvons ajouter des "marqueurs de sujet" : John-sujet aime Mary-objet. Mais maintenant, nous avons besoin de différents neurones (ou clusters) pour représenter le sujet-John et l'objet-John. Et que dire de "Jean aime Marie et Susan aime Jean et Marie aime Susan" (chaque personne est à la fois sujet et objet mais cela ne suffit pas..)

# FODOR ET PYLYSHYN

- En résumé, le grand défi (de Fodor et Pylyshyn) pour les connexionnistes est le suivant : devons-nous nier que l'intelligence nécessite une véritable systématique, devons-nous trouver une nouvelle forme de systématique qui ne nécessite pas de structure partie-entière, ou devons-nous trouver une forme connexionniste de structure partie-entière (qui ne met pas simplement en œuvre un système classique de suivi des règles) ?

# FODOR ET PYLYSHYN

- En arrière-plan, il faut toutefois savoir que lorsque Fodor et Pylyshyn écrivait, les modèles connexionnistes des processus clés n'étaient pas performants et il n'existait pas d'algorithme d'apprentissage efficace. (On soupçonnait donc qu'aucune alternative utile au raisonnement syntactico-symbolique ne fonctionnerait jamais, même pour la classification complexe des chiens par rapport aux chats...)