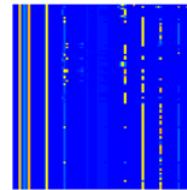
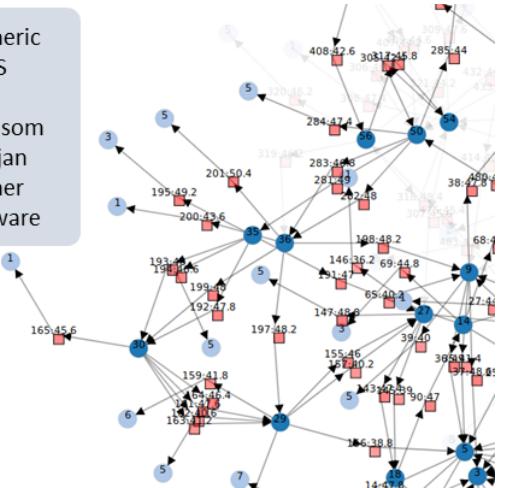


« System calls »
seen by AI



- 1 : Generic
- 2 : DOS
- 3 : Spy
- 4 : Ransom
- 5 : Trojan
- 6 : Miner
- 7 : Adware



CBWAR

Classification de Binaires Windows via Apprentissage par Renforcement
(Windows Binary Classification using Reinforcement Learning)

Olivier Gesny

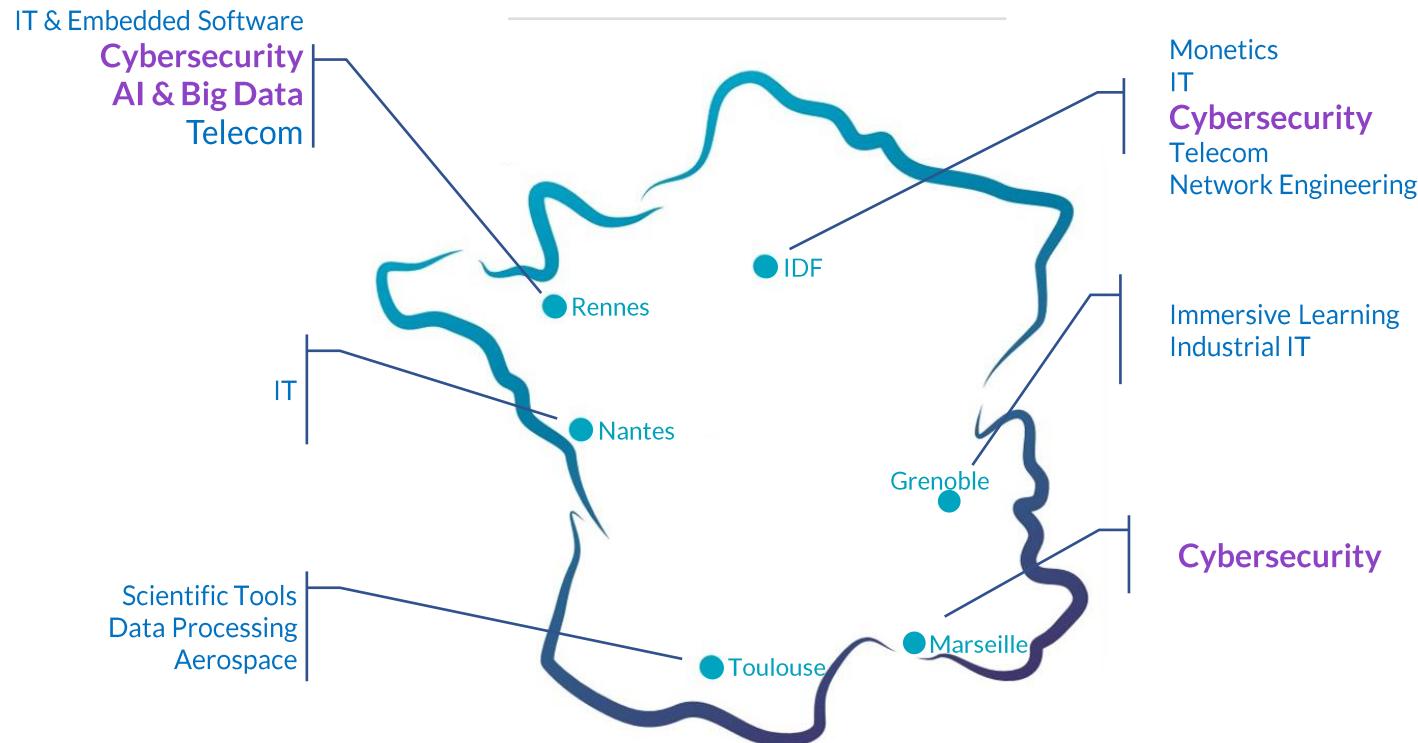
Responsable Pôle Embarqué, Cybersécurité et IA
Head of team Embedded, Cybersecurity and AI

Silicom.®

4, rue de Jouanet - 35700 - Rennes
+33 (0) 2.99.84.17.17
ogesny@silicom.fr
www.silicom.fr

Silicom.

at the heart of a community of experts



www.silicom.fr

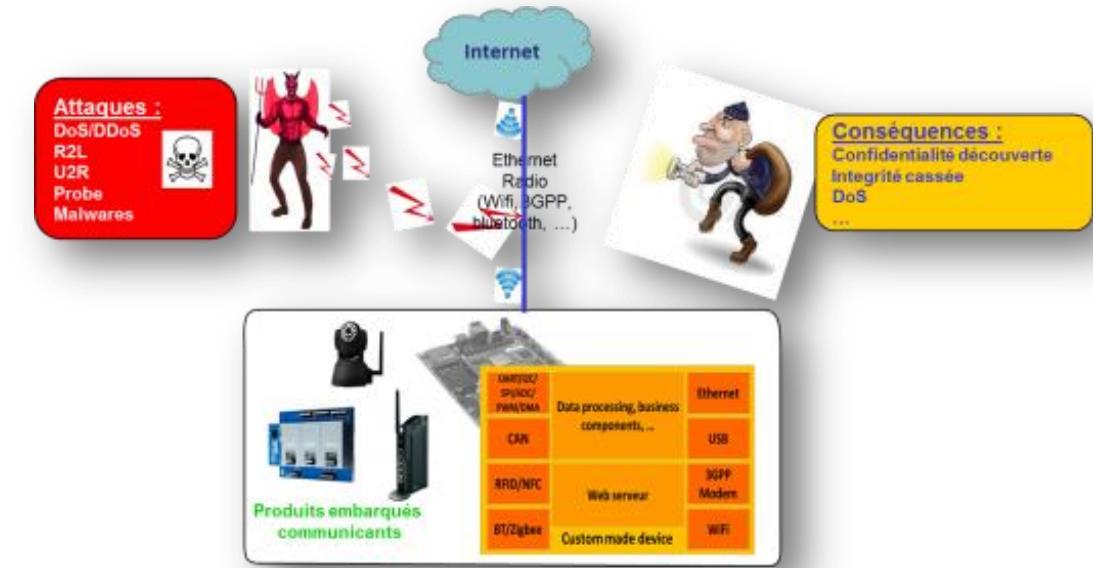
Motivation du projet

Au cœur de notre stratégie : la cybersécurité et l'IA

- Nous développons des systèmes complexes et cybersécurisés
- Nous développons des systèmes de tests automatiques (fonctionnels et de découverte de vulnérabilités) pour garantir la sûreté des logiciels
- Nous menons des projets d'exploitation de données massives (Big Data)

Des attaques qui se multiplient et qui deviennent de plus en plus pernicieuses

- MIRAI : attaque DDOS sur Dyn DNS à partir de caméras infectées
- StuxNet : attaque des centrifugeuses nucléaires Iraniennes
- Ransomware WannaCry, Petya, NotPetya, Ryuk (2016, 2019) : attaque des systèmes informatiques de très nombreuses entreprises publiques ou privées
- Images ou pdf malveillants qui exploitent des failles de sécurité logicielles (windows, acrobat, viewer d'images, etc)
- Andromeda/Gamarue (2011 à 2017) : botnet gérant une famille de 80 malwares
- APT (Advanced Persistent Threat) : passent inaperçues pendant une période très longue (plusieurs mois à plusieurs années)



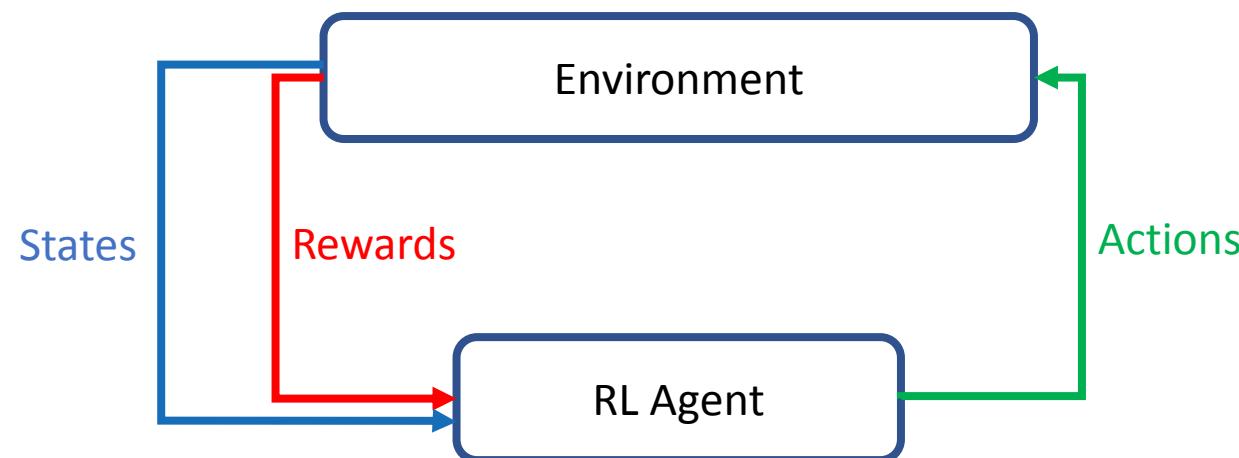
Nous souhaitions appliquer des techniques d'IA récentes à la détection d'attaques



Concept et vocabulaire

Reinforcement Learning (ou Apprentissage par Renforcement)

- Un [Agent] évolue dans un environnement [environment] dans lequel il effectue des actions
- Pour chaque action qu'il entreprend, l'agent :
 - constate des changements d'états [states] de l'environnement
 - et récupère des récompenses (appelées [rewards])
- L'objectif pour l'agent est de trouver la meilleure stratégie d'actions (appelée [policy]), celle qui conduit à maximiser les récompenses
- Le Reinforcement Learning est dans ses fondements un algorithme semi-supervisé (dans le sens où il a besoin de données labellisées par l'humain)
- Avec le Reinforcement Learning, un agent apprend au travers d'essais, réussites et erreurs. Il repose donc sur des mécanismes que l'on considère proches du comportement humain





Quelques succès retentissants (1/6)

DQN - DeepMind (01/2015)





Quelques succès retentissants (2/6)

Deal or no deal? Training AI bots to negotiate – Facebook AI Research (06/2017)

The screenshot shows a web-based interface for a negotiation task. On the left, a sidebar displays instructions and rules for the negotiation. On the right, a main panel shows the items available for division and a message input field for communication with a partner.

Divide these items between you and your partner.

Your partner sees the same items but with different values

You get some items, and your partner will get the rest

If you get a great deal for you then we will pay a bonus!

If you often get low scores then your work may be rejected

Items to Split between You and Partner	Value Each to You	Number You Get
2 Books	0	0 <input type="button" value=""/>
1 Hat	7	0 <input type="button" value=""/>
3 Basketball	1	1 <input type="button" value=""/>

Deal was Agreed!

Fellow Turker connected. Please send a message!

Type Message Here:

Message

No deal was agreed



Quelques succès retentissants (3/6)

Emergence of locomotion behaviours in rich environment – Google Deepmind (07/2017)





Quelques succès retentissants (4/6)

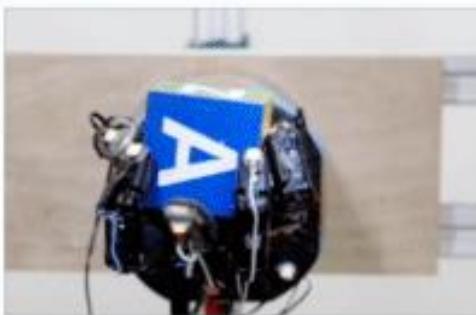
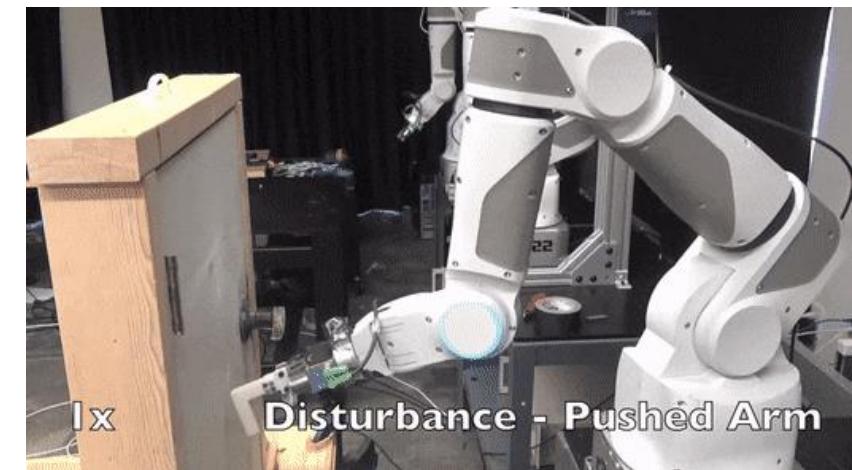
OpenAI five – OpenAI (06/2018)





Quelques succès retentissants (5/6)

Robotique



FINGER PIVOTING



SLIDING



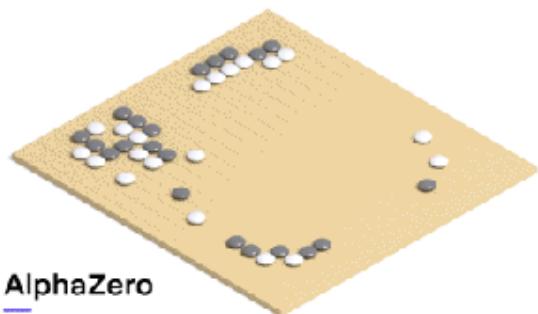
FINGER GAITING

Learning Dexterity – OpenAI (07/2018)

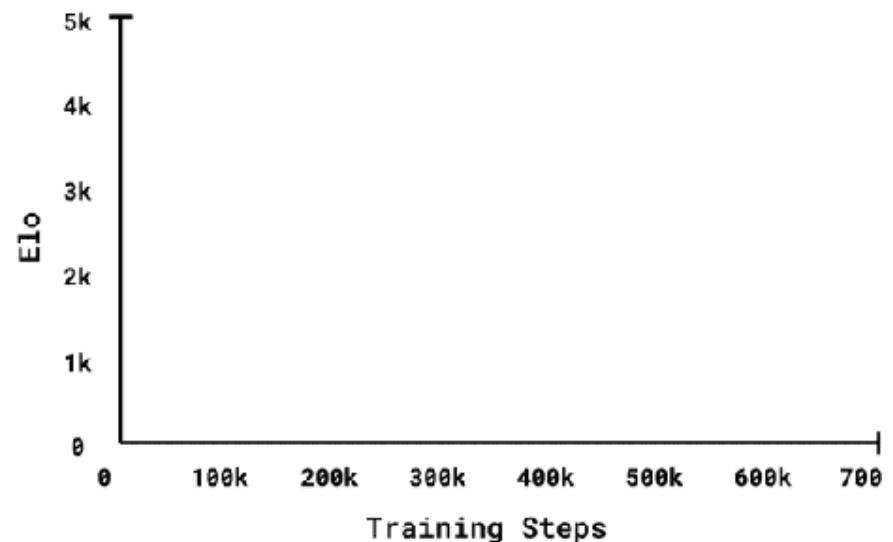


Quelques succès retentissants (6/6)

AlphaGo, AlphaGoZero et AlphaZero - DeepMind (01/2016, 05/2017 et 12/2017)



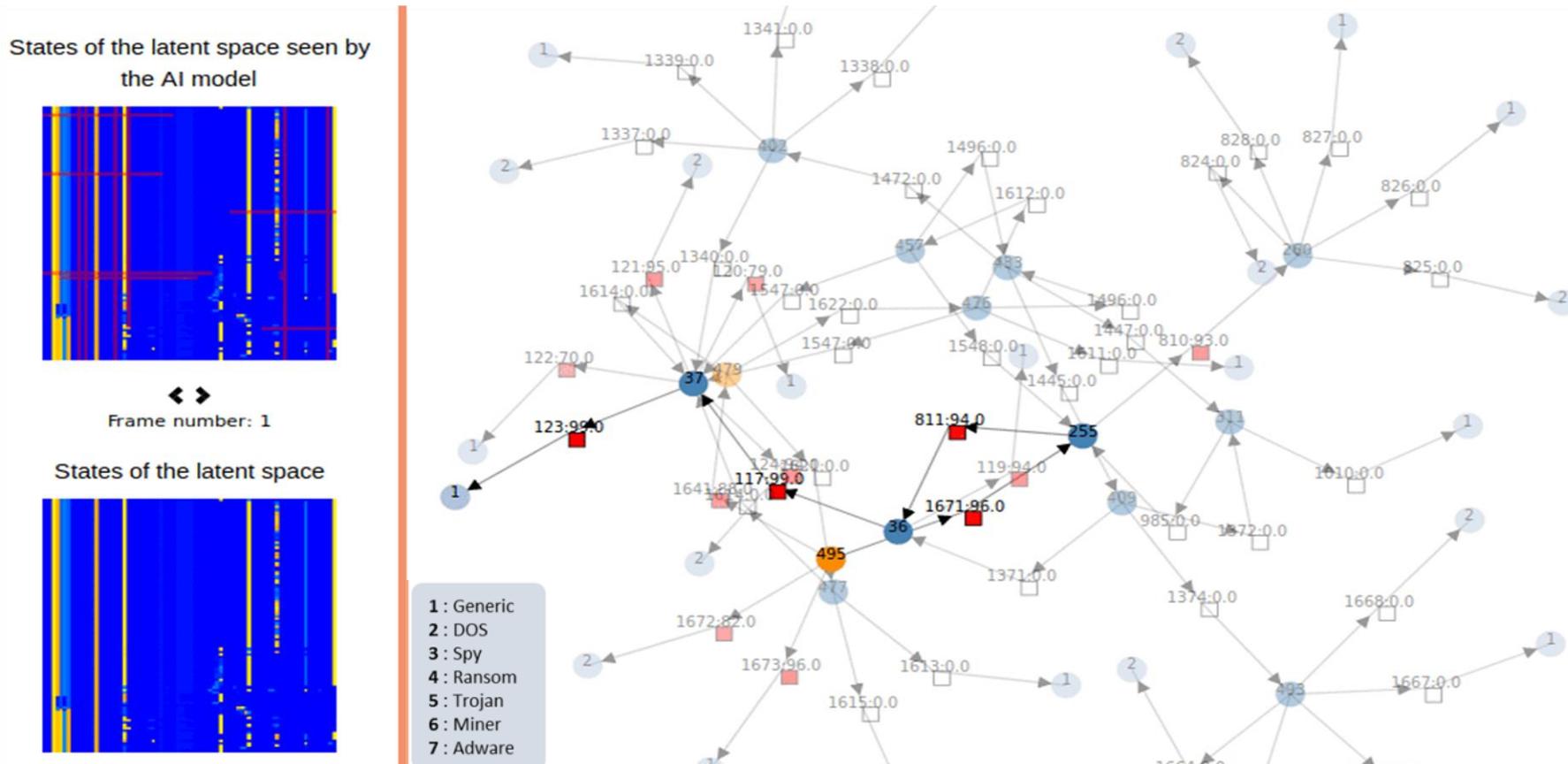
AlphaZero





Notre contribution : SIVA

Cyberdéfense : Classification multiclass de binaires Windows (2018)





Travaux relatifs à notre contribution

TPG (Tangled Program Graphs)

- Meilleur papier à l'Euro GP 2017
- Meilleur papier à l'ACM GECCO 2017



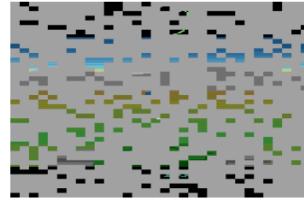
Ms. Pac-Man Screen



Ms. Pac-Man AVF



Battle Zone Screen



Battle Zone AVF

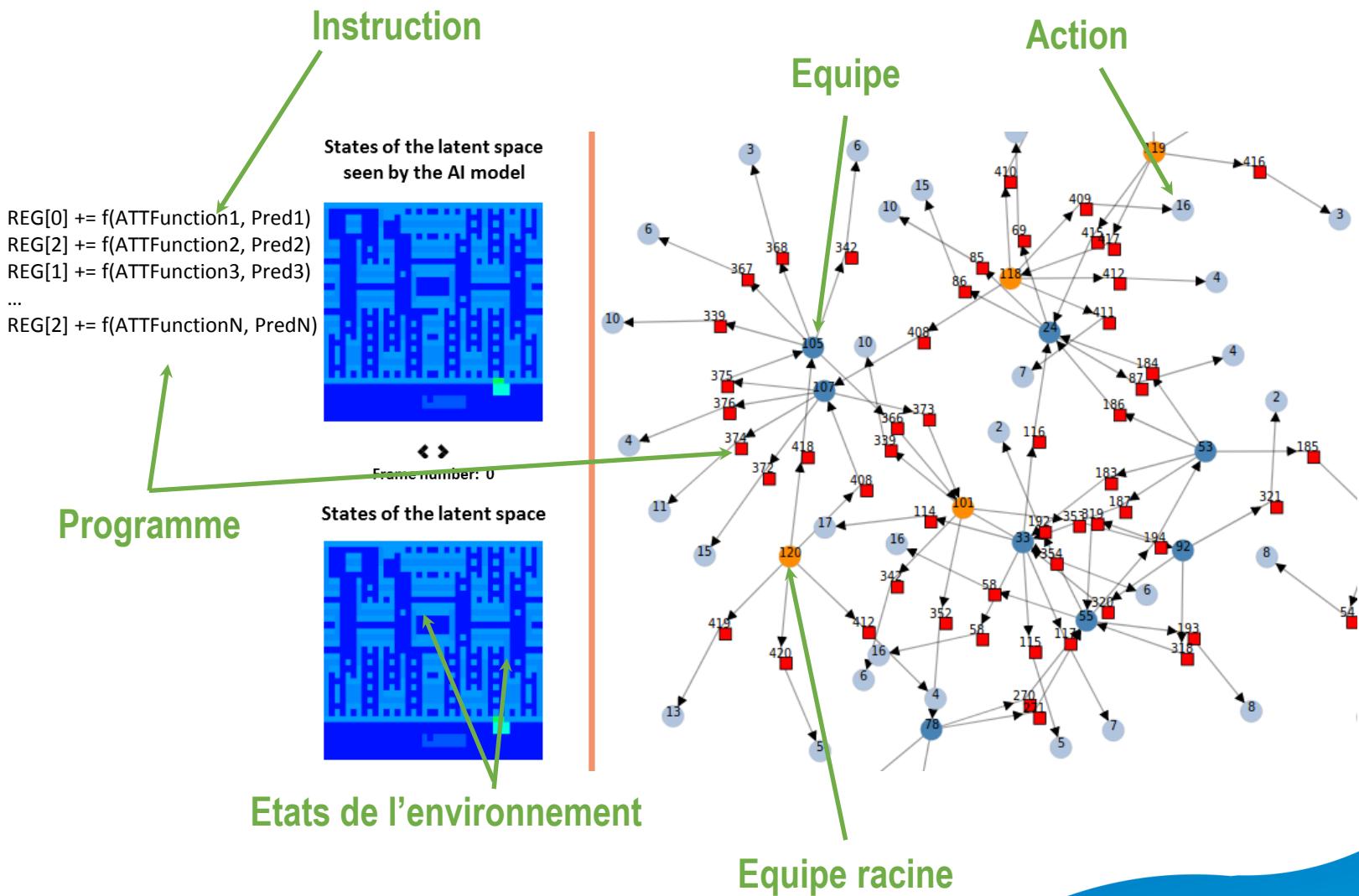
Game	DQN	HNEAT	Hum	TPG	Tms	Ins	%IP
Alien	3069(± 1093)	1586	6875	3382.7 (± 1364)	46	455	56
Amidar	739.5 (± 3024)	184.4	1676	398.4(± 91)	63	812	69
Asterix	6012 (± 1744)	2340	8503	2400(± 505)	42	414	51
Asteroids	1629(± 542)	1694	13157	3050.7 (± 947)	13	346	23
BankHeist	429.7(± 650)	214	734.4	1051 (± 56)	58	572	65
BattleZone	26300(± 7725)	36200	37800	47233.4 (± 11924)	4	123	11
Bowling	42.4(± 88)	135.8	154.8	223.7 (± 1)	56	585	57
Centipede	8309(± 5237)	25275.2	11963	34731.7 (± 12333)	28	516	39
C.Command	6687(± 2916)	3960	9882	7010 (± 2861)	51	280	58
DoubleDunk	-18.1(± 2.6)	2	-15.5	2(± 0)	4	116	6
Frostbite	328.3(± 250.5)	2260	4335	8144.4 (± 1213)	21	382	28
Gravitar	306.7(± 223.9)	370	2672	786.7 (± 503)	13	496	36
M'sRevenge	0	0	4367	0(± 0)	18	55	28
Ms.Pac-Man	2311(± 525)	3408	15693	5156 (± 1089)	111	1036	83
PrivateEye	1788(± 5473)	10747.4	69571	15028.3 (± 24)	59	938	60
RiverRaid	8316 (± 1049)	2616	13513	3884.7(± 566)	67	660	64
Seaquest	5286 (± 1310)	716	20182	1368(± 443)	22	392	37
Venture	380(± 238.6)	NA	1188	576.7 (± 192)	3	165	7
WizardOfWor	3393(± 2019)	3360	4757	5196.7 (± 2550)	17	247	31
Zaxxon	4977(± 1235)	3000	9173	6233.4 (± 1018)	20	424	33



Concepts du module IA (SIVA)

SIVA (Silicom Versatile AI) overview

- La politique Reinforcement Learning est représentée par un graphe
- Des **équipes racine** d'**agents/programmes** évoluent et se concurrencent dans un environnement
- Chaque **équipe racine** constitue le nœud d'entrée d'un arbre de décision
- A chaque génération, toute **équipe racine** est en compétition contre les autres équipes
- Des **équipes standards (sub-team)** et des **agents/programmes** peuvent travailler pour plusieurs **équipes racine**
- Les **agents/programmes** sont constitués d'instructions exécutées par l'IA pour prendre une **décision d'action**
- Les meilleures **équipes racines** sont renforcées : elles se reproduisent, mutent et continuent à évoluer alors que les autres sont éliminées



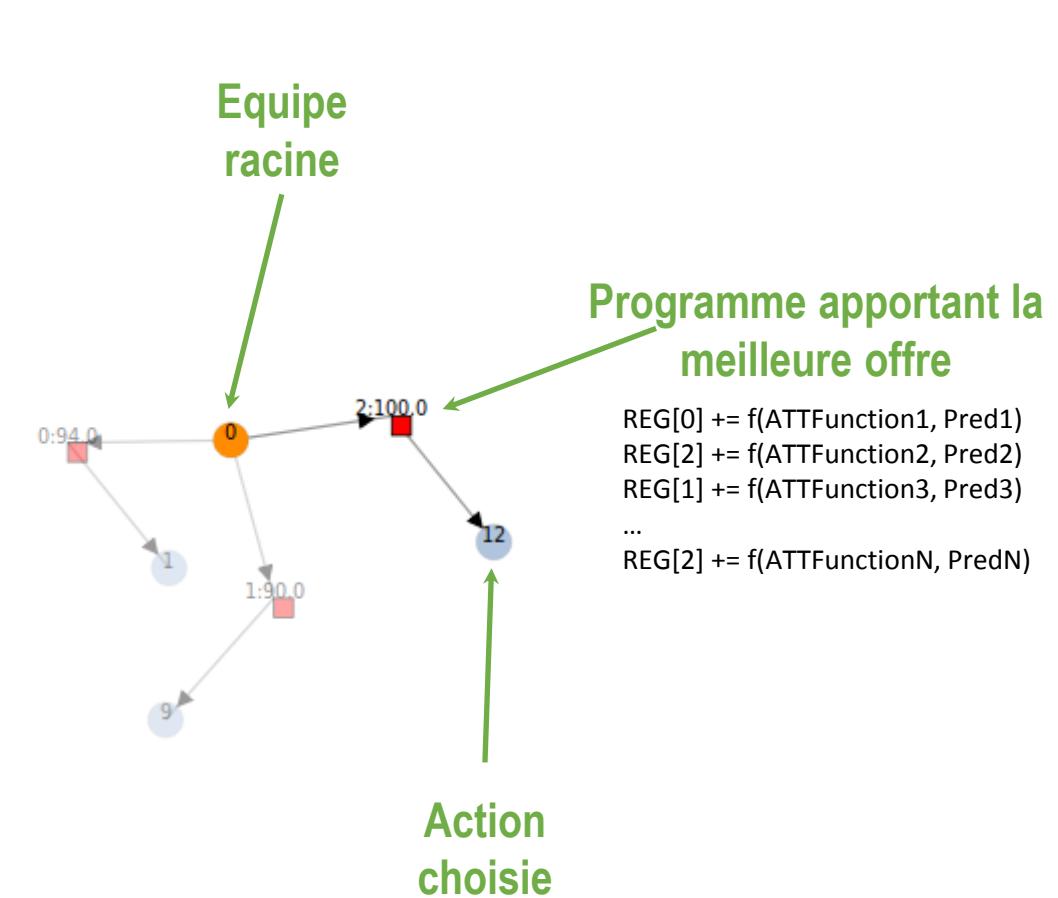
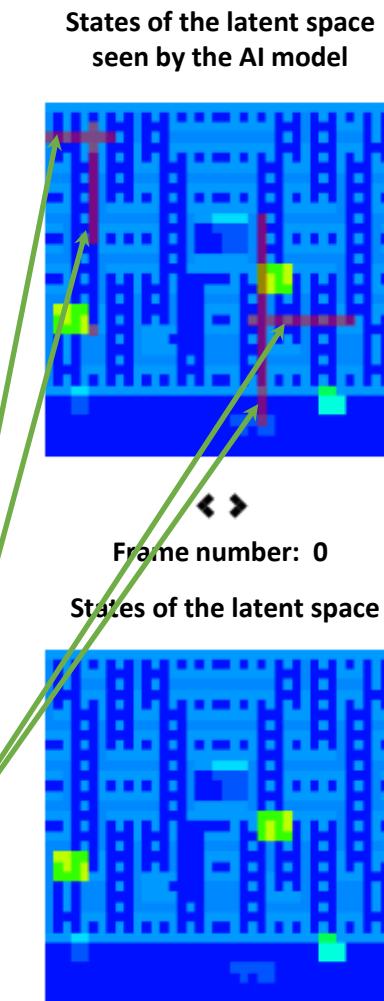


Concepts du module IA (SIVA)

Prise de décision SIVA pendant l'apprentissage

- Principe d'observation/perception (fonction d'attention) comparée à une prédition
- Chaque **programme**:
 - Exécute ses instructions (qui prédisent et observent des propriétés)
 - Calcule une offre basée sur l'observation de la propriété courante/passée et la qualité de prédition définie pour chaque instruction
 - Pointe vers une **décision d'action** ou une autre équipe de programmes
- Réalisation de l'**action**

Adresses/états de l'environnement observées par l'équipe racine



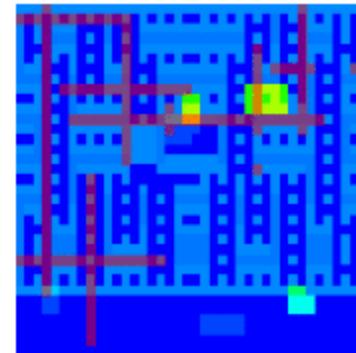


Concepts du module IA (SIVA)

Prise de décision SIVA pour une équipe racine

- Chaque **programme** :
 - Exécute ses instructions (qui prédisent et observent des propriétés)
 - Calcule une offre basée sur l'observation de la propriété courante/passée et la qualité de prédiction définie pour chaque instruction
 - Chemin vers une **action** finale ou une autre **équipe** de **programmes**
 - **Décision d'action**

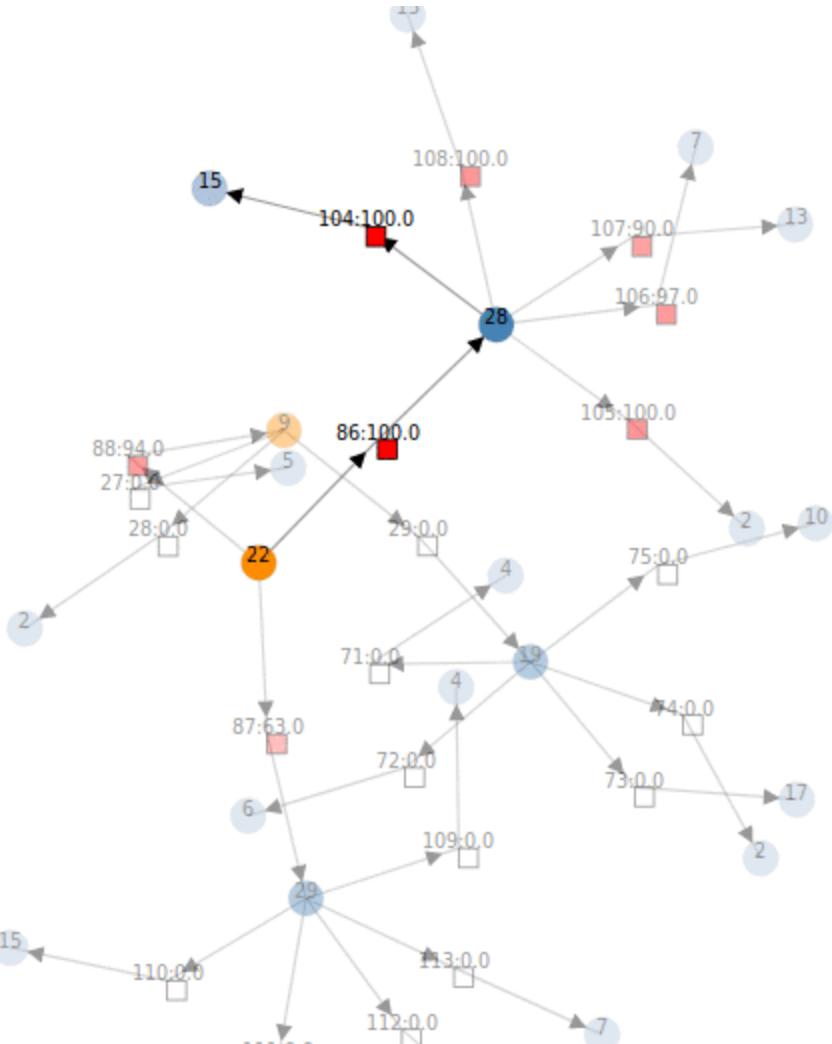
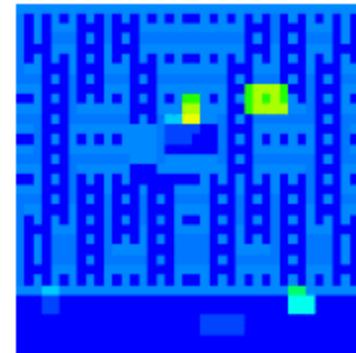
States of the latent space seen by the AI model



2

Frame number: 0

States of the latent space

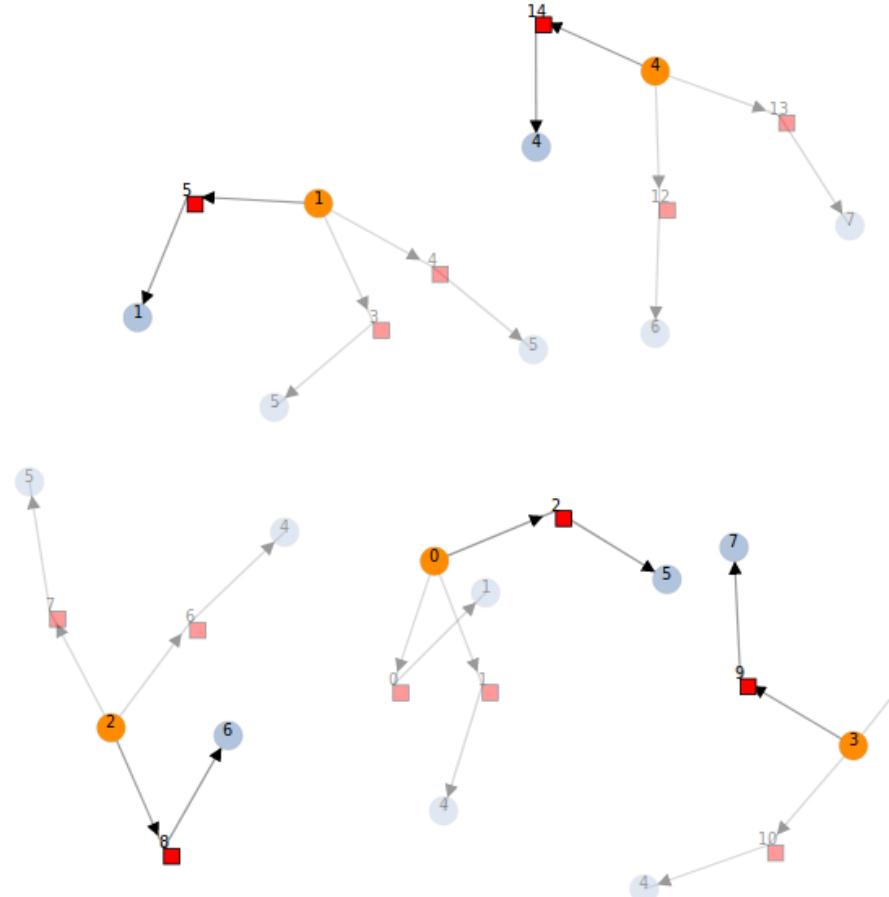




Concepts du module IA (SIVA)

Evolution de SIVA en cours d'apprentissage

- N équipes racine en compétition
- Toute équipe racine exécute, chacune à son tour, un batch de challenges du dataset d'entraînement
- Après consolidation des récompenses obtenues, les pires équipes sont éliminées
- La meilleure équipe est clonée puis mutée

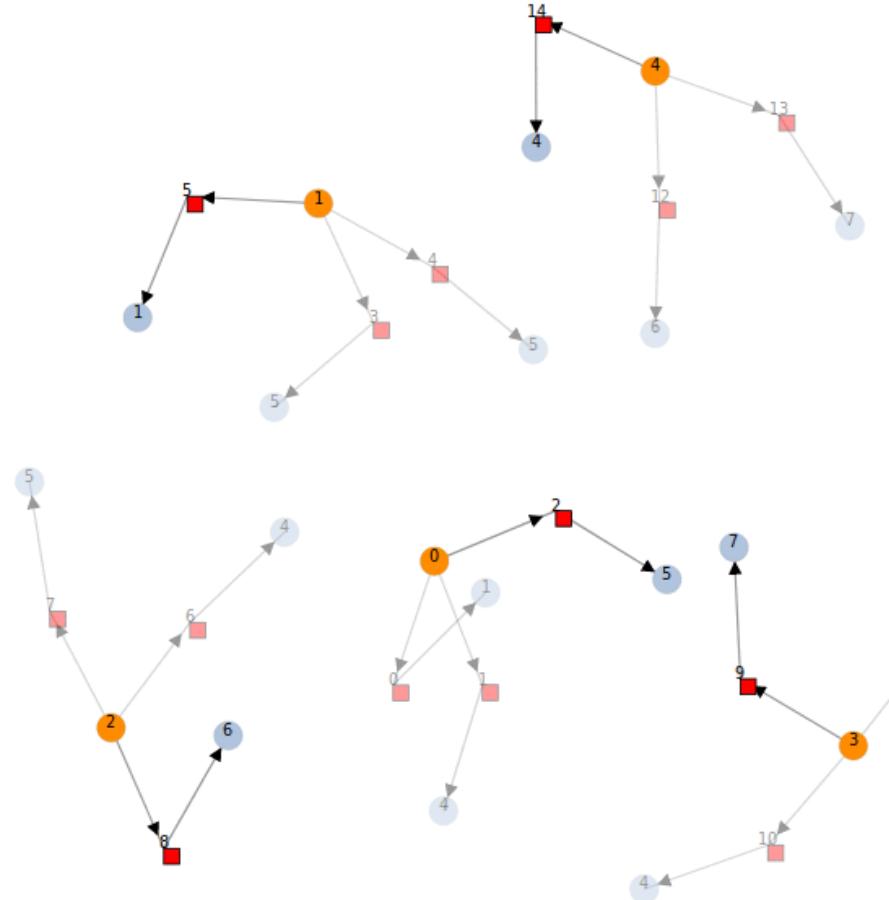




Concepts du module IA (SIVA)

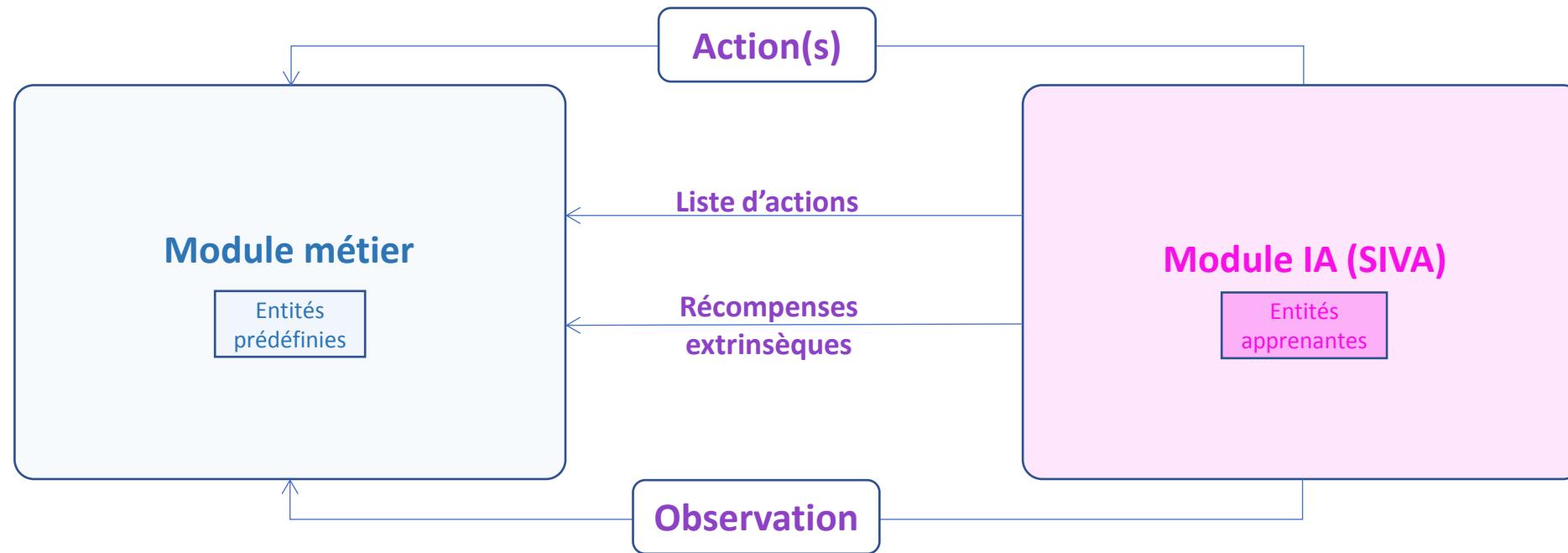
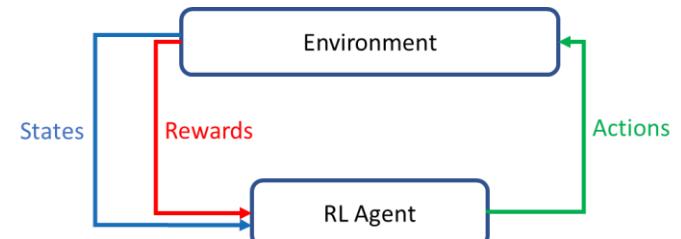
Evolution de SIVA en cours d'apprentissage

- N équipes racine en compétition
- Toute équipe racine exécute, chacune à son tour, un batch de challenges du dataset d'entraînement
- Après consolidation des récompenses obtenues, les pires équipes sont éliminées
- La meilleure équipe est clonée puis mutée





Architecture globale d'un modèle

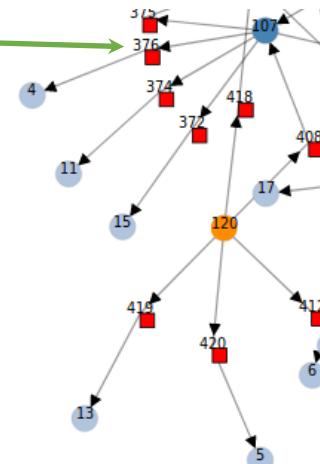




Description d'un modèle – Focus sur l'instruction

Programme

```
REG[0] += f(ATTFunction1, Pred1)  
REG[2] += f(ATTFunction2, Pred2)  
REG[1] += f(ATTFunction3, Pred3)  
...  
REG[2] += f(ATTFunctionN, PredN)
```



Composition d'une instruction (prédiction et calcul d'une propriété de l'environnement)

- **tpgRegister**: registre TPG
- **operation**: intitulé de la fonction d'attention (distances, états, transitions temporelles, nombre/surface)
- **currentRefAddressInEnvironment**: adresse de l'"objet référence" dans l'environnement présent
- **currentCompAddressInEnvironment**: adresse de l'"objet comparé" dans l'environnement présent
- **elapsedDurationSinceLatestObservation**: saut temporel écoulé depuis la dernière observation de l'"objet "
- **foveaSize**: rayon d'observation de la fonction d'attention
- **currentRefState**: état courant à l'adresse currentRefAddressInEnvironment
- **currentCompState**: état courant à l'adresse currentCompAddressInEnvironment
- **passedRefState**: état passé de l'adresse currentRefAddressInEnvironment à $t - \text{elapsedDurationSinceLatestObservation}$
- **passedCompState**: état passé de l'adresse currentCompAddressInEnvironment à $t - \text{elapsedDurationSinceLatestObservation}$
- **distanceRefComp**: distance entre les 2 objets comparés
- **predictedProperty**: propriété prédite
- **acceptablePercentageOfErrorToExactProperty**: pourcentage d'erreur acceptable pour la propriété prédite au regard de la propriété exacte
- **numberOfAcceptablePredictions**: nombre de prédictions acceptables
- **meanInstructionPredictionQuality**: moyenne de la qualité de prédiction de l'instruction. Mise à jour exclusivement lorsque la prédiction est acceptable
- **numberOfExecutions**: nombre d'exécutions de l'instruction



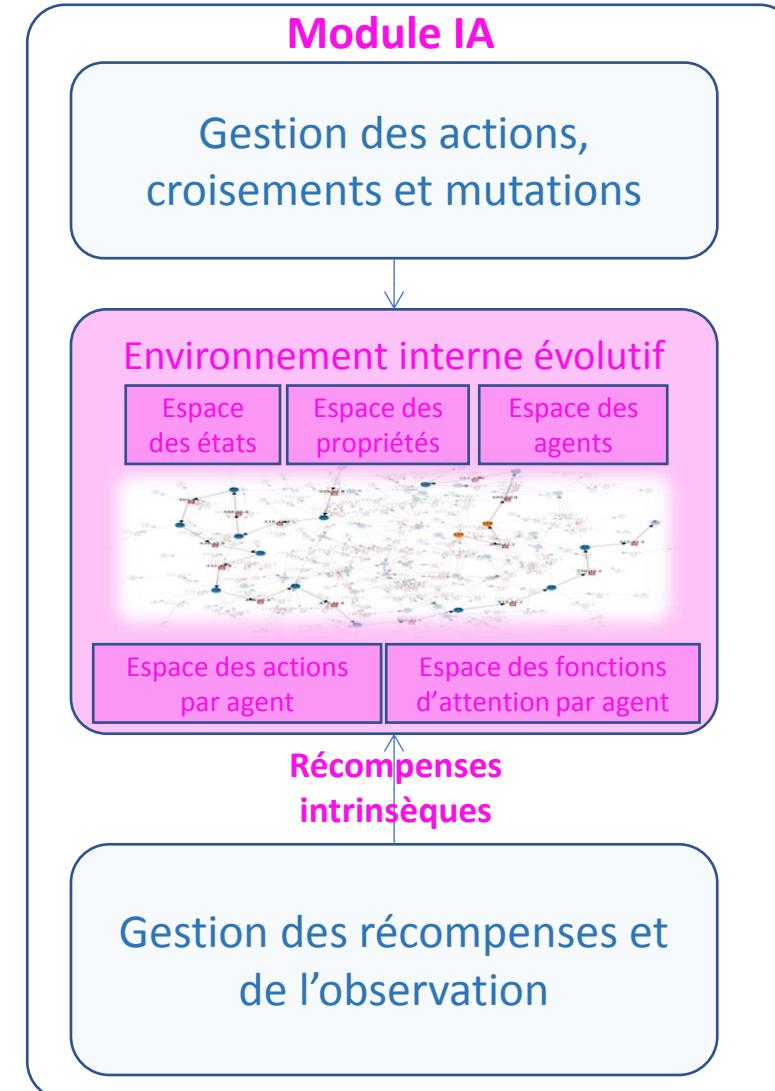
Description d'un modèle – Description du module AI (SIVA)

Principales caractéristiques

- Environnement interne évolutif
- Enrichissement des programmes et instructions au regard du TPG
 - Fonctions d'attention (distances, états, transitions temporelles, nombre/surface)
 - Calcul de la qualité de prédiction instantanée (représente l'écart entre la propriété prédite et la propriété exacte)
 - Correction des instructions en live
 - Mutation des instructions en live
- Le programme ancre toute nouvelle prédiction presque correcte

Exemples d'innovations

- Agents prédictifs : prédiction des propriétés associées à un objet unique ou à 2 objets
- Récompenses intrinsèques instantanées pour chaque instruction dont la qualité de prédiction est dans un range acceptable
- L'offre d'un programme est la contribution des récompenses intrinsèques (fonction des qualités de prédiction) obtenues pour chaque instruction



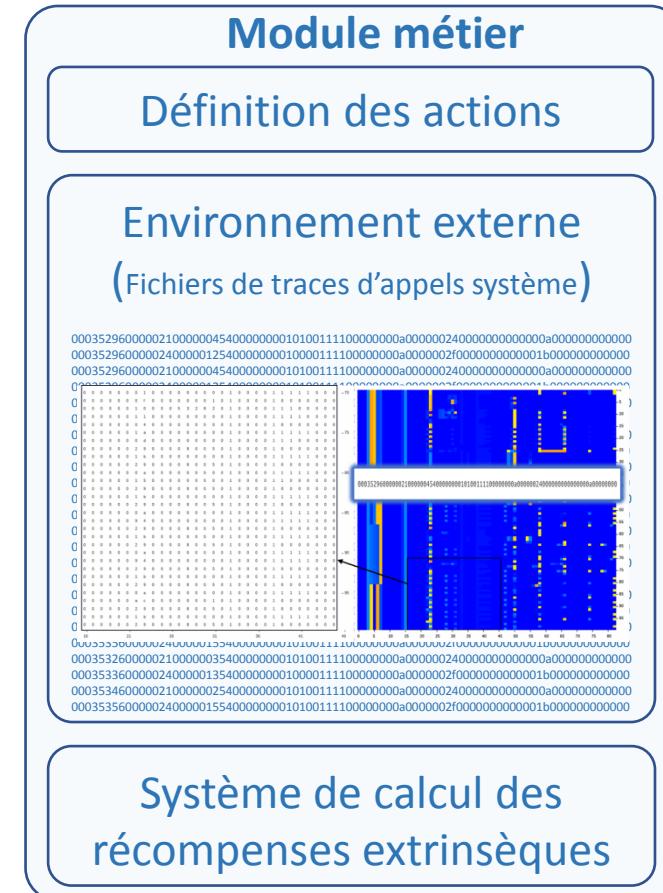


Description d'un modèle – Description du module métier (ex : CBWAR)

Principales caractéristiques

- Construction de l'environnement à partir de l'environnement réel (traces d'appels système)
- Définition des actions unitaires
 - Classification
 - Déplacement (ex : lire les 33 lignes suivantes, aller à droite, à gauche, etc)
 - Tout autre action (exécution script, appel API)
- Structuration du système de récompense
 - Calcul d'indicateurs métier (ex : FPR, TPR)
 - Définition d'indicateurs symboliques de l'embodiment (positif ou négatif)
 - La calcul du score dépend d'objectif métier, des indicateurs calculés et des signaux d'embodiment

$$Score = f(TP \cdot \text{PondTP}; TN \cdot \text{PondTN}; FP \cdot \text{PondFP}; FN \cdot \text{PondFN})$$

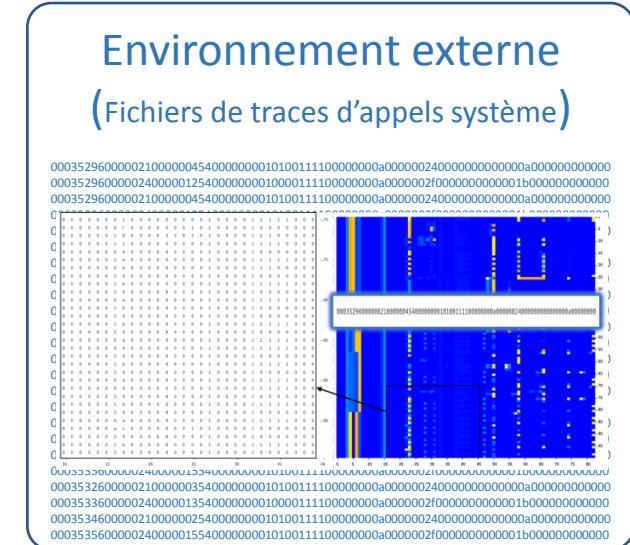




Expérimentations – Focus sur l'environnement observé

Constitution de l'environnement 'Appels système'

- 100 lignes de 83 octets
- Format (octets) d'une ligne 'Appels système'
 - [0-7] -> time_ms
 - [8-15] -> time_s
 - [16-23] -> id
 - [24-31] -> ret_value
 - [32-33] -> succes
 - [34-35] -> interesting_function
 - [36-37] -> interesting_argument
 - [38-42] -> arg_used
 - [43-82] -> arguments des appels système (5 arguments pour CBWAR mais aucune limite pour le modèle)



Dataset

- Ensemble de fichiers d'appels système json de tous les binaires Windows PE à notre disposition : extraits de rapports de la sandbox cuckoo
- Labellisation Virus Total pour chaque binaire ({SAFE, GENERIC, DOS, SPY, RANSOM, TROJAN, MINER, ADWARE})
- Dataset d'entraînement : 2773 fichiers json
- Dataset d'évaluation : 2157 fichiers json
- Moyenne de 9579 appels système par binaire

```

        "tags": [
            "desired_access": ""
        }
    },
    {
        "category": "synchronisation",
        "status": 1,
        "stacktrace": [],
        "api": "NtCreateMutant",
        "return_value": 0,
        "arguments": [
            "initial_owner": 1,
            "desired_access": "0x001f0001",
            "mutant_name": "KyUffTh0kYwRRtgPP",
            "mutant_handle": "0x00000084"
        ],
        "time": 1527414453.07775,
        "tid": 2796,
        "flags": [
            "desired_access": "STANDARD_RIGHTS_ALL|STANDARD_RIGHTS_REQUIRED|DELETE"
        ]
    },
    {
        "category": "system",
        "status": 1,
        "tags": [
            "desired_access": "0x00000004",
            "return_value": 0
        ]
    }
],
{
    "0": "NtCreateMutant",
    "1": "00035296",
    "2": "00000021",
    "3": "00000045",
    "4": "40000000",
    "5": "01",
    "6": "01",
    "7": "00",
    "8": "11110",
    "9": "0000000a",
    "10": "00000024",
    "11": "00000000",
    "12": "0000000a",
    "13": "00000000"
},
{
    "0": "RegOpenKeyExW",
    "1": "00035296"
}

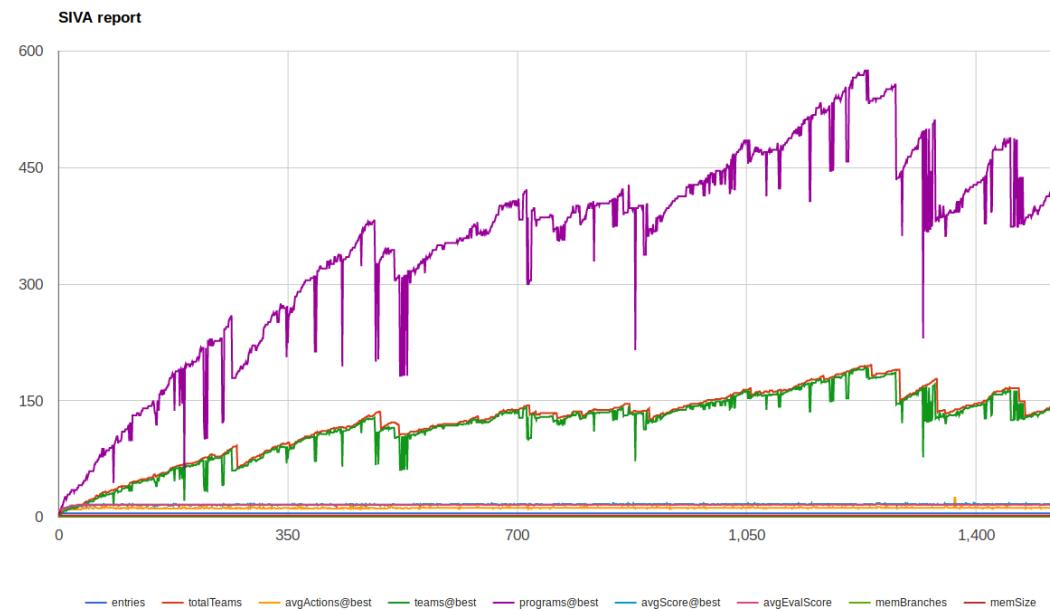
```





Résultats

Expérimentation	Durée d'apprentissage (ms)	FPR (%)	ACC (%)	TPR (%)	TNR (%)	FNR (%)
Fenêtre d'acquisition : 100 appels système Pondérations pour convergence vers FPR faibles : - TP : 0.75; FP : 1.0; TN : 0.0; FN : 0.0	113478 (~32h)	0,28	89,16	71,67	99,76	27,12



$$Score = \frac{TP \cdot \text{PondTP} + TN \cdot \text{PondTN}}{FP \cdot \text{PondFP} + FN \cdot \text{PondFN}} \text{ avec } FP \cdot \text{PondFP} + FN \cdot \text{PondFN} > 0$$



Conclusion

Exemples de plus-values

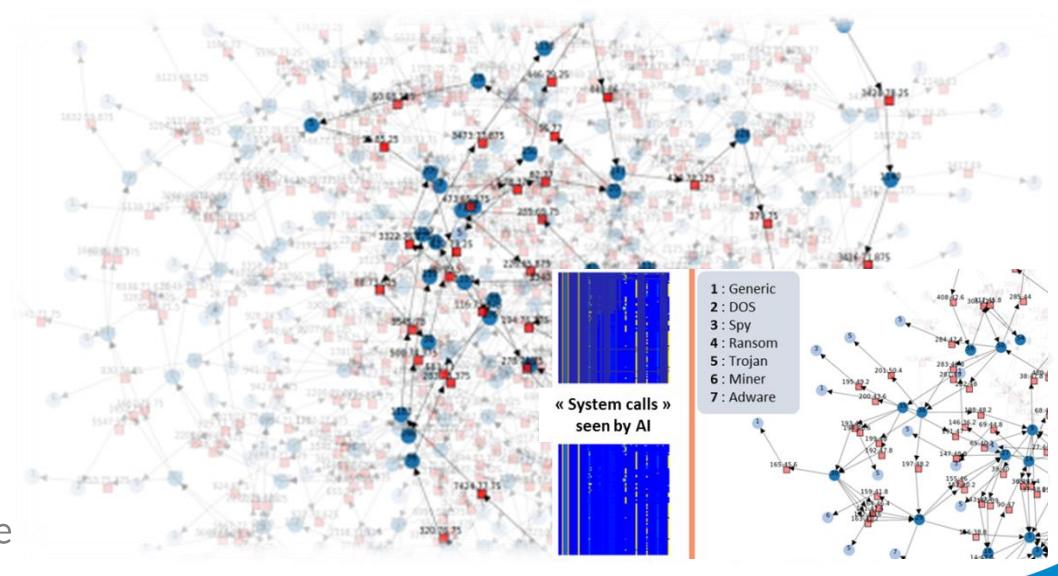
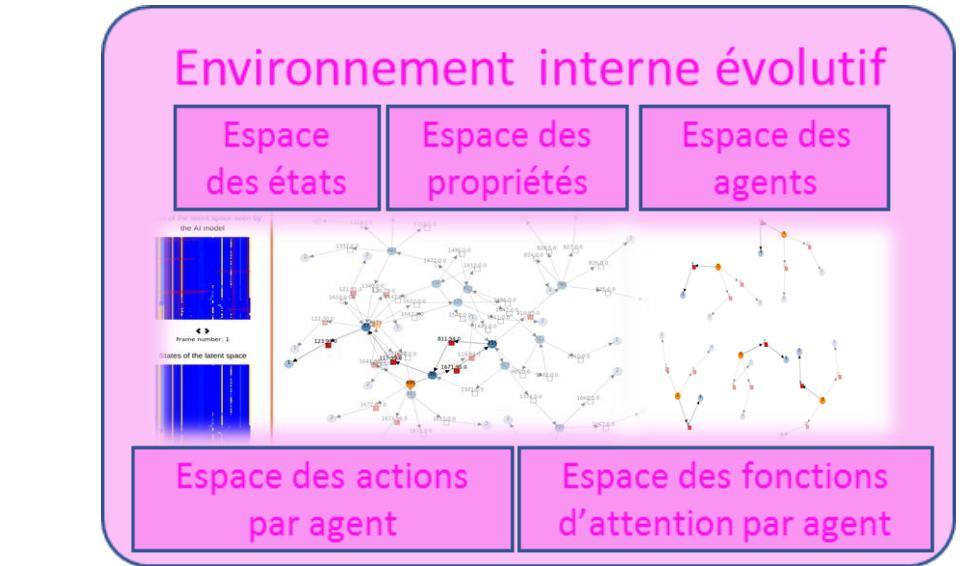
- Emergence d'un arbre de décision (decision tree)
- Fortes corrélations entre fonctions d'attention (perception des agents), actions et objectif à atteindre grâce à l'émergence de liens pertinents entre les équipes, les programmes et les actions
- Explicabilité
- Pouvoir de segmentation/discrimination
- Adaptable à de nombreux cas d'usage (times series mining, classification, détection de comportements anormaux, optimisation, etc)

Faiblesses

- Les mutations sont encore trop aléatoires donc les liens pertinents entre agents n'émergent pas aussi rapidement qu'ils pourraient
- Code non parallélisé
- L'apprentissage n'est pas aussi rapide qu'il pourrait l'être

Attention aux considérations éthiques

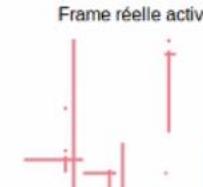
- Définir avec soin les objectifs en respectant les aspects éthiques
- Définir avec soin les actions unitaires : certaines d'entre elles sont dangereuses si elles sont données initialement au modèle apprenant => Ne mettez pas une arme dans les mains d'un bébé





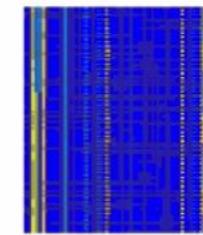
Vidéo de démonstration

```
• ScU r0 [6891] #{41}  
#(48) #(48) #"0" #"-43"  
#<1> #|1| e1 p1 a66  
c100 b1 k3  
• SdD r0 [6891] #{41}  
#(48) #(52) #"-4" #"1"  
#<1> #|1| e100 p1 a65  
c600 b6 k3  
• SdD r0 [6891] #{41}  
#(48) #(53) #"-5" #"2"  
#<0> #|1| e100 p1 a36  
c700 b7 k3  
• Scl r0 [6891] #{42} #(48)  
#(54) #"-6" #"0" #<1>  
#|1| e1 p1 a55 c200 b2 k3  
• StD r0 [6891] {41} #(48)  
#(53) #"-5" #"0" #<1>  
#|16| e100 p0 a26 c700  
b7 k2  
• StU r0 [6891] {42} #(48)  
#(50) #"-2" #"0" #<1>  
#|3| e100 p0 a58 c700 b7  
k2  
• StL r0 [6891] {42} #(48)  
#(50) #"-2" #"0" #<0>  
#|2| e100 p0 a67 c600 b6  
k2  
• SdL r0 [6891] #{42}  
#(48) #(51) #"-3" #"1"  
#<0> #|1| e1 p1 a60  
c700 b7 k3  
• StD r0 [6891] {42} #(48)  
#(51) #"-3" #"0" #<1>  
#|3| e100 p0 a60 c800 b8  
k2  
• SdL r0 [6891] #{42}  
#(48) #(48) #"0" #"1"  
#<0> #|1| e1 p1 a19  
c700 b7 k3  
• ScU r0 [6891] #{42}  
#(48) #(49) #"-1" #"-39"  
#<0> #|1| e1 p1 a36
```



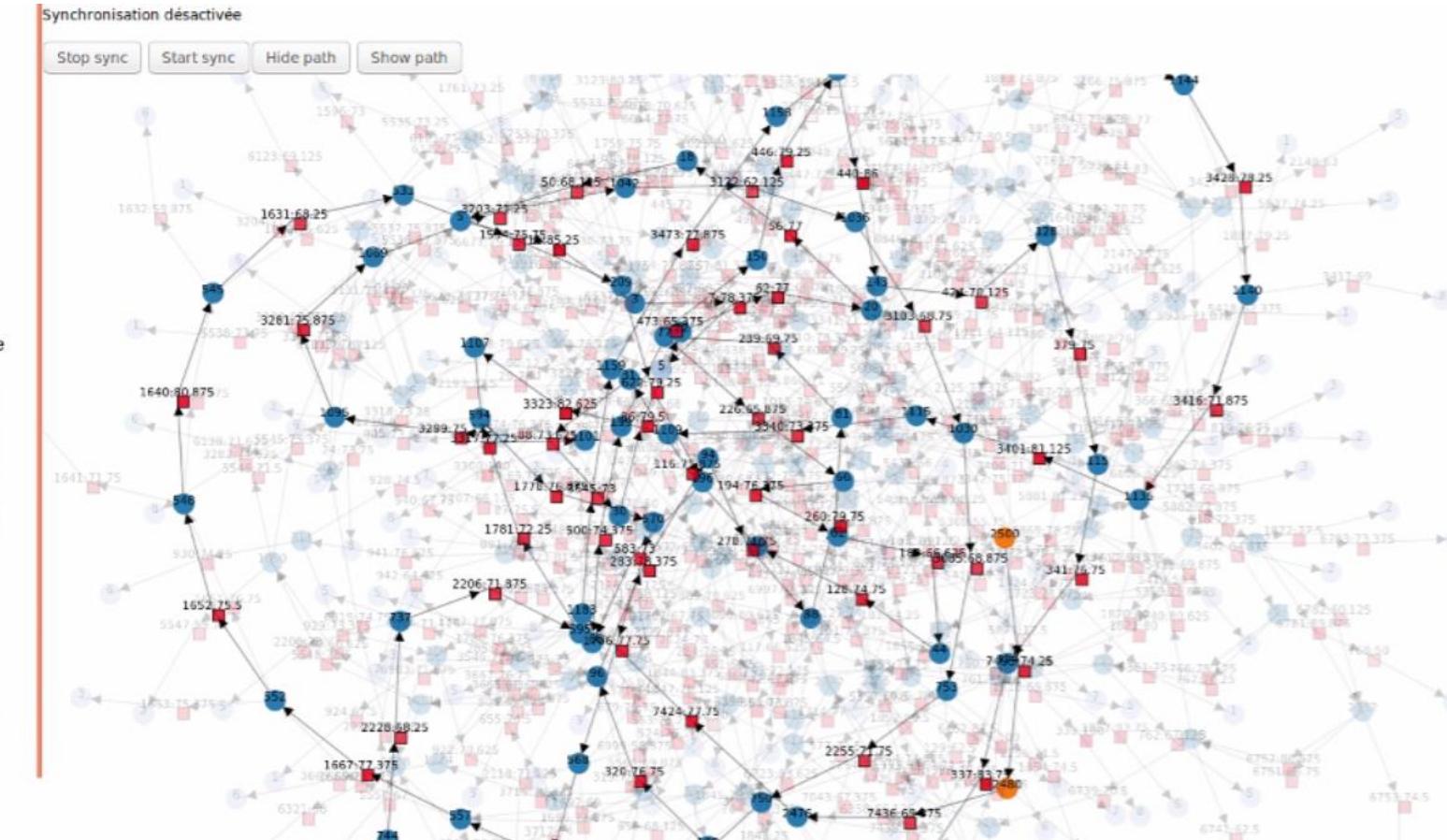
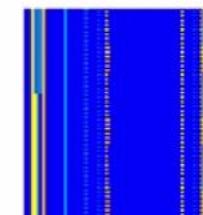
Frame réelle active
Frame numéro : 24

Frames vue par l'agent dans le stimulus



Frame numéro : 0

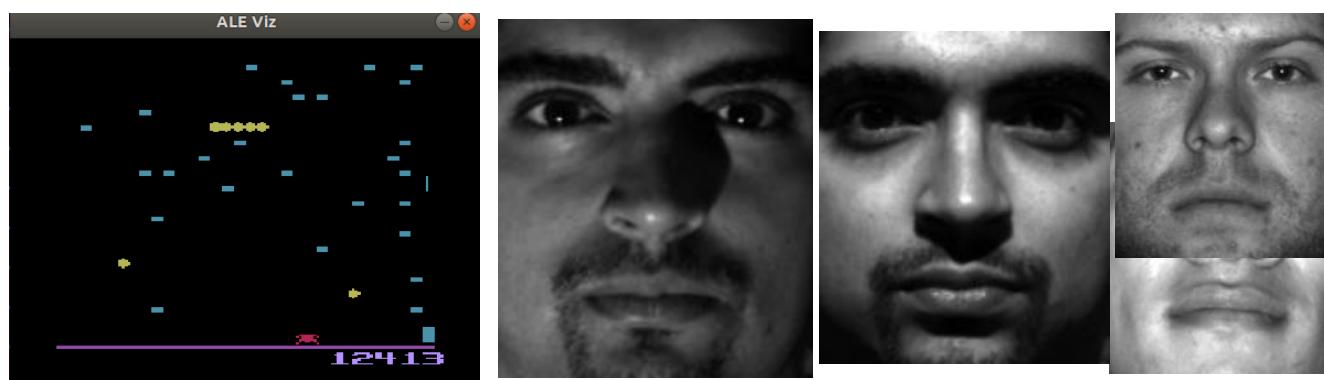
Valeurs du stimulus





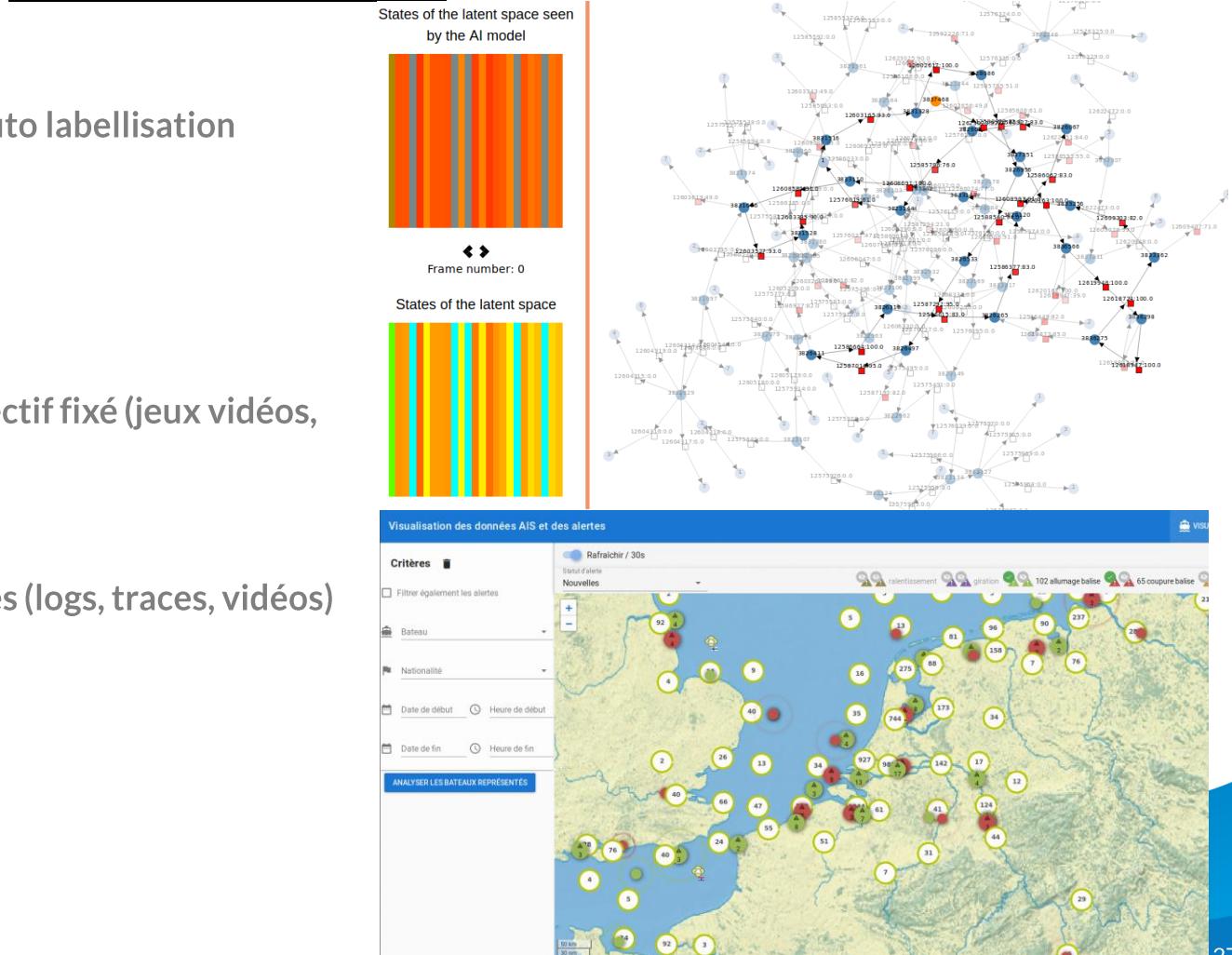
Amélioration de SIVA

- Multisensor/multimodal
- Prédiction d'objets de taille plus importante
- Récompenses intrinsèques supplémentaires
- Apprentissage à partir de datasets de taille réduite et auto labellisation
- Mutations des liens entre agents/programmes



Cybersécurité, détection/classification et planification

- Détection/classification de fichiers (pdf, images, etc)
- Planification des actions permettant d'atteindre un objectif fixé (jeux vidéos, reconnaissance faciale)
- Détection d'intrusion dans des SI (thèse à venir)
- Détection/Classification de binaires malveillants
- Recherche de signaux faibles dans des séries temporelles (logs, traces, vidéos)
- Détection de comportements atypiques ou anormaux



L'évolution du vivant est une source d'inspiration pour améliorer notre IA

- Comportement des organismes et leur comportement
- Evolution symbiotique
- Auto organisation
- Neurosciences
- Cellules apprenantes (Learning cells)



<https://www.popularmechanics.com/science/a23842/slime-mold-slime-lapse/>

La curiosité est une source d'inspiration pour améliorer notre IA

- Oudeyer, Lecun, Bayes, Pearl, Domingo
 - Lachaux, Dehaene, Dessalles, Houdé, Sablonnière, Changeux
 - Karsenti, Bapteste, Dussutour
 - Berthoz, Debru, Jeannerod, Damier
 - Ameisen, Klein, Lê Nguyên Hoang

PHYSARUM POLYCHILOPHUM, ou blob, n'est ni un animal, ni une plante, ni un champignon. C'est un organisme unicellulaire et supraspécifique qui présente des propriétés de l'apprentissage et peut se déplacer de quelques centimètres par heure, tout en faisant 4 directions.

► Bien que dépourvu de neurones, il est capable d'apprendre et de se souvenir, à traverser un pent court couvert de subtils répulsifs alors qu'il avance.

► Plus étonnant, un blob peut transmettre à un autre blob sa capacité à traverser des pentes.

LES AUTEURS



AUDREY DUSSUTOUR
chercheuse au CNRS au Centre de recherche sur la cognition animale, à l'Ecole



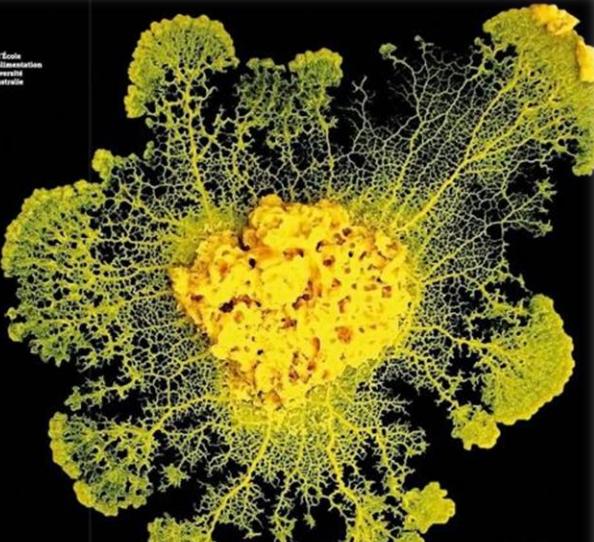
DAVID VOGEL
professeur à l'Ecole d'agriculture, d'alimentation et de pêche de l'université d'Amérique du Sud, à Béziers

Le blob, cellule géante... et intelligente!

ETONNANT ORGANISME MACROSCOPIQUE CONSTITUÉ D'UNE SEULE CELLULE, LE BLOB PEUT APPRENDRE ET MÊME TRANSMETTRE SON SAVOIR. IL EST CAPABLE DE TROUVER LE PLUS COURT CHEMIN POUR SORTIR D'UN LABYRINTHE OU D'IMITER LE RÉSEAU FERROVIAIRE JAPONAIS.

La diversité des formes de vie sur Terre dépasse l'imagination et ne cesse de nous surprendre. Dans ce fascinant monde, les cellules sont les plus communes, mais bien d'un point de vue physiologique que comportemental. Ainsi, nous connaissons tous les organes unicellulaires, mais aussi l'ensemble des plantes, d'apprentissage des hums ou encore de construction des fourmis. D'ailleurs, les capacités des organismes unicellulaires sont étonnantes, voire étonnantes.

Pourtant, les organismes unicellulaires présentent des comportements qui n'ont rien à voir avec la vie pluricellulaire. Par exemple, certaines sont capables de communiquer, de s'orienter, de se déplacer, de construire des abris... Un exemple étonnant est celui des



découvertes récemment chez un organisme unicellulaire de forme et de taille indéfinies, visqueux et de couleur jaune vif, le myxomycète *Physarum polycephalum*.

UN ORGANISME INCLASSABLE

Le place de *Physarum polycephalum* dans l'arbre des végétaux a longtemps été incertaine. Il a été classé à diverses périodes soit comme plante (à cause de ses feuilles), soit comme animal (à cause de sa couleur et de sa forme), soit comme champignon (à cause de ses spores). En 1753, le naturaliste suédois Carl von Linné a classé les myxomycètes parmi les plantes, en s'appuyant sur leur façon de se déplacer des plantes. Mais au contraire de ces dernières - diversité qui rappelle celle des plantes, puisqu'en trouvant des myxomycètes (plus d'un million d'espèces connues

© M. Sano et al. 2010, Springer-Verlag Berlin Heidelberg

Le blob fait son tour sur le monde d'un peu des années 1950. Mais d'après ce que nous savons maintenant, il est néanmoins quelqu'un de très intelligent. Il existe au moins 47 000 espèces de myxomycètes dans le monde, mais il n'a été décrit qu'en plusieurs milliers.

Audrey Dussutour & David Vogel - Pour la Science N°483 - Janvier 2018