

# Créons l'Observatoire des Évolutions Algorithmiques !

Le concept de singularité technologique [1] modifie profondément notre perception du progrès et de l'évolution humaine.

Simple dogme ? prévision éclairée ? ou techno-prophétie arrogante ?

A vrai dire, peu importe l'idée que l'on s'en fait car nous vivons chaque jour les événements prémices de cette singularité. Il suffit de les observer côte à côte, sur un même plan de lecture pour distinguer nettement la diffusion algorithmique qui vient bouleverser des marchés, des pratiques professionnelles et des existences.

## Chroniques du Code

Quel lien transparent peut bien unir un trader des marchés financiers basé à Londres, spécialisé dans le négoce des produits gaziers, un journaliste sportif à Miami, assurant depuis plus de vingt ans les commentaires en direct de match de Base-ball, un activiste s'apprêtant à franchir clandestinement la frontière sud-coréenne, ou encore un tailleur de saphir travaillant sur une très belle pierre brute à Jaipur ?

L'algorithme bien sûr ! Chacun des protagonistes va entrer en compétition, parfois sous la forme d'une concurrence très déloyale, avec un algorithme.

Le trader qui ne pratique pas encore d'algo-trading (ou HFT pour High Frequency Trading) a appris à connaître (à ses dépens) et à redouter les assauts des codes invisibles et fulgurants qui agissent et déforment le marché en quelques dixièmes de secondes.

Après trois ou quatre prises de positions perdantes, il finira par éviter certaines plages horaires du marché qu'il sait risquées car privilégiées par les codes HFT.

On fera lire au journaliste le compte-rendu du dernier match entre les Florida Marlins et les Chicago Cubs, il y verra un article très classique, fidèle à la rencontre, sans originalité particulière. Il apprendra aussitôt qu'aucun de ses collègues n'est à l'origine de l'article mais qu'il est né d'un algo-rédacteur Stats Monkey acheté par sa rédaction chez Narrative Science il y a deux semaines, et que l'article en question coûte moins de 10 dollars.

Le journaliste, fragilisé par cette froide concurrence électronique, refusera longtemps d'entendre le pari d'un des concepteurs de Stats Monkey selon lequel, d'ici à cinq ans, un algo-rédacteur remportera le prix Pulitzer...

Dès ses premiers pas en zone militaire sécurisée, notre clandestin sera pris en charge par une des nombreuses bornes de surveillance offensive Samsung SGRA1 positionnées le long de la frontière nord de Corée du sud.

La borne analysera le comportement de notre imprudent randonneur, calculera le niveau de menace et prendra seule la décision de neutraliser l'intrus en ouvrant le feu : pas d'intervention humaine ici, mais seulement un algorithme décidant d'un tir de 12mm.

Enfin, notre lapidaire indien, devra se faire violence et accepter la stratégie de taille déterminée par un code subtil mêlant géométrie, optimisation et modélisation.

Le brut de saphir aura été scruté, simulé, reconstruit par le programme.

Ses impuretés prises en compte, le calcul décidera de la meilleure découpe.

L'opération de taille pourra alors débiter, sereinement, passant outre le millénaire d'expérience humaine des lapidaires indiens.

Il nous est facile de poursuivre cette « chronique du code » et d'imaginer ce que seront les rencontres algorithmiques de demain : en dérivant la fonction de l'algo-rédacteur, on visualise son utilisation lors de débats politiques ou économiques réunissant experts, personnalités politiques et animateurs journalistes : le dispositif est alors complété par un algo-modérateur qui analyse en direct les questions, les réponses, les arguments (chiffrés ou non) avancés par chacun et qui intervient lorsqu'il décèle une contre-vérité, un reniement d'idée, une hypothèse contradictoire ou une statistique erronée. On imagine sans peine les changements à apporter au discours politique et

la nécessité d'inventer une nouvelle langue de bois, forcément algo-compatible...

La même classe d'algorithme peut être envisagée lors d'un conseil d'administration, d'une réunion contradictoire, d'une cellule de crise au sein d'un ministère, d'une structure militaire ou encore dans le cadre d'une délibération de justice en cour d'assises.

L'entité algorithmique contrôlera en temps réel l'avancée des discussions, l'exactitude des données, la pertinence des déductions, et calculera les conséquences des décisions potentielles.

Elle sera à la fois un expert, un contrôleur, un modérateur et à terme, un prévisionniste via la construction de simulations.

Ces pratiques aujourd'hui émergentes vont bouleverser de façon disruptive les prises de décisions humaines.

Comment peut-on se préparer à ces mutations brutales et à leurs effets collatéraux de type « cygnes noirs » [2] ?

Faut-il emboîter le pas américain avec la création d'une grande structure de formation, de recherche, catalyseur de convergence, comme l'université de la singularité ? [3]

En avons-nous les moyens et la volonté ?

**Dans la suite de l'article, nous allons nous montrer optimiste et téméraire afin d'imaginer ce que peut être un dispositif minimal de préparation à la singularité.**

**Partons de l'idée d'une structure basique de veille technologique ciblée sur la sphère algorithmique.**

**Commençons par la nommer : OEA pour Observatoire des Évolutions Algorithmiques.**

**Fixons lui ses premières missions, son architecture et sa localisation :**

## **Missions**

1- L'OEA assure une veille technologique appliquée au champ algorithmique.

Il élabore une cartographie détaillée, en temps réel, des algorithmes actifs, recense puis analyse les innovations émergentes en partenariat avec laboratoires universitaires, sociétés de développement de logiciels et GENCI (Grand Équipement National de Calcul Intensif).

Il évalue, à l'aide de critères standardisés, les indices de qualité, d'adaptabilité, de potentiel stratégique et/ou économique associés à un algorithme identifié.

2 - L'OEA favorise le transfert technologique : il opère en direction des acteurs de l'informatique française : cluster, sociétés de service, web-village, pme-pmi, banque-trading, sociétés de conseil et d'expertise, administrations et ministères.

Toute entreprise ou administration développant un projet à forte composante algorithmique peut saisir l'OEA afin de commander une analyse exhaustive fixant le champ algorithmique lié au projet. L'étude établit alors un panorama comparatif de l'existant algorithmique lié à la demande, elle produit ensuite une analyse optimisée des technologies à mettre en œuvre ou des adaptations à apporter à une entité algorithmique disponible.

3 - L'OEA est en mesure d'initier une enquête d'algo-profilage à la suite d'un événement de nature algorithmique sans que l'algorithme ne soit initialement identifié et décrit (seuls ses effets sont accessibles). La cellule mobilisée mène alors une investigation en vue d'identifier ou de caractériser le groupe d'algorithme à l'origine de la perturbation constatée; l'OEA procède à la manière d'un « bureau enquête-accident » et identifie d'éventuelles interactions ou instabilités algorithmiques.

4 - L'OEA se prononce sur la compatibilité d'utilisation d'un algorithme avec le cadre législatif en vigueur : l'algorithme A est-il légalement conforme ? Ce « certificat » peut être délivré en collaboration avec l'actuelle CNIL.

L'observatoire développe une infrastructure d'analyse d'objet numérique, procède à une

authentification d'origine de production et délivre un certificat.

Concrètement, l'analyse de l'objet précise s'il est d'origine humaine selon l'indice de confiance p , s'il est d'origine mixte, c'est-à-dire humaine au niveau h et algorithmique au niveau 1-h avec l'indice de confiance p ou bien totalement algorithmique.

L'objet en question peut être un texte, un livre, une image, un son, une vidéo ou un objet physique. Cette mesure d'origine (qui s'apparente à un test de Turing) s'avérera stratégique en terme de cyber-sécurité, d'intelligence économique ou militaire.

5 - L'OEA mesure le caractère « disruptif » et stratégique d'un nouvel algorithme. Il évalue la hauteur d'impact d'une utilisation généralisée de l'algorithme sur son environnement d'application. Cette évaluation peut être appuyée par toute forme d'enquête, de test (comme le test de Turing) ou de simulations numériques.

6 - L'OEA milite auprès des organismes de formation (grandes écoles, instituts, universités) pour une prise en compte globale de l'outil algorithmique comme vecteur de convergence des sphères de connaissances classiques (neurosciences, biotique, calcul haute performance, sciences cognitives, sciences sociales, finance, marketing, ...).

7 - L'OEA se positionne comme un catalyseur de la convergence; il réalise un maillage stratégique, via un réseau de correspondants, en intégrant chaque composante participant au mouvement de convergence et en veillant à maintenir une approche transversale du progrès algorithmique.

## **Architecture**

L'architecture d'une structure influence son efficacité : il faut imaginer ici une entité dynamique, évolutive et cellulaire autorisant une adaptation rapide aux évolutions disruptives.

Une cellule centrale supervisant l'ensemble du dispositif est construite autour d'un comité scientifique.

Cette cellule alpha peut se développer au sein de structures existantes à profils compatibles :

- 1 - Centre d'analyse stratégique dépendant du Premier Ministre.
- 2 - Institut de recherche stratégique de l'école militaire , IRSEM ministère de la défense.
- 3 - Ministère de l'économie numérique ou futur « Paris Numérique -web village ».
- 4 - Centre de Calcul Recherche et Technologie , CCRT , CEA, GENCI.
- 5 - Fondation pour la recherche stratégique – FRS (structure indépendante).

Il s'agit d'une liste non exhaustive, car d'autres structures ou fondations sont en mesure d'accueillir notre cellule alpha.

La composition du conseil scientifique doit refléter l'image d'une future convergence, elle doit en épouser les formes.

Dressons une liste non exhaustive de 13 profils indispensables :

- Un membre expert GENCI-CEA.
- Un membre expert DGA.
- Un membre expert biotechnologie – biotique.
- Un membre expert neurosciences.
- Un membre expert sciences cognitives - psychologie cognitive.
- Un membre expert finance-trading.
- Un membre expert web-business – économie numérique.
- Un membre expert sociologie et philosophie des changements.

- Un membre expert édition de logiciels – services et conseils informatique.
- Un membre expert marketing.
- Un membre expert médias.
- Un membre expert théorie de l'information.
- Un membre expert cyber-sécurité - hygiène informatique - renseignement.

Il convient de compléter le groupe par un représentant du ou des ministères de tutelle de l'OEA.

Notre cellule alpha constituée, peut à présent débiter une phase classique de réseau-ification en tissant une trame multi-domaines de correspondants OEA :

Laboratoires de recherche, CNRS, CEA, Universités, Grandes Ecoles, Départements R. et D.industriels., Banques-assurances, cluster, futures structures de type web-village régionaux, Agences de développement,.. la liste est évidemment ouverte et évolutive.

Chaque correspondant validé par la cellule alpha constitue une nouvelle cellule de la structure. L'information et l'innovation algorithmique doivent pouvoir diffuser rapidement et librement au sein de cette trame, ceci constitue une condition nécessaire d'efficacité.

## **Dynamique**

Illustrons maintenant notre construction par quelques cas concrets de dynamique OEA :

Un groupe bancaire opérateur de trading HFT saisit l'OEA et sollicite la construction d'une simulation numérique de type multi-agents afin d'étudier les risques associés aux flash crash HFT. Le simulateur devra reconstituer une zone du marché, faire interagir trader humains, algo-traders, et organe de régulation. Les tests permettront une analyse des crashes émergents et des instabilités. L'OEA facilite et accélère la construction du simulateur en mobilisant les cellules concernées (GENCI, Université, et/ou société de développement) et en participant à la conception de l'architecture globale du projet.

Une administration est victime d'une attaque numérique visant son système d'information. L'OEA peut alors effectuer une étude de rétro-analyse, de reconstruction et de profilage des algorithmes agressifs mis en œuvre ; l'observatoire peut ensuite proposer une stratégie de défense, en activant les cellules pertinentes.

L'OEA encourage et soutient la création de sociétés d'édition de logiciels antivirus Français ; et milite en faveur du maintien d'un niveau minimal de souveraineté dans ce domaine sensible. Concrètement, l'OEA supporte des projets liés à la sécurité comme DAVFI antivirus open source français.

Enfin, l'OEA peut mettre en place un environnement de test « à la Turing » ciblant un nouvel algorithme de rédaction ou d'expertise crée par une société d'édition de logiciels. Les tests externalisés évitent ainsi les biais d'évaluation souvent présent lors « d'un test maison » .

L'OEA intervient au sein des universités et grandes écoles au moyen de conférences transversales décrivant certaines classes d'algorithmes et les futurs besoins de développements associés. L'observatoire met l'accent sur les territoires algorithmiques en devenir et dresse un bilan transversal de l'existant.

Pour conclure, décrivons ce que peut contenir un profilage algorithmique :

## Profil algorithmique

Nous noterons  $A$  un algorithme ou  $A = (A_1, A_2, \dots, A_N)$  un groupe d'algorithmes et distinguerons deux situations d'approche.

**Situation 1 :** L'algorithme  $A$  est disponible, nous avons accès à sa description précise et complète dans un langage formel  $L$ .

Concrètement,  $A$  est codé par un programme  $P$ , nous disposons du code source, c'est-à-dire de toute l'information liée à  $A$ .

**Situation 2 :** L'algorithme  $A$  n'est pas disponible, nous ne percevons que ses effets sur un environnement  $E$  lui aussi partiellement accessible.

Cette situation est de complexité supérieure à la précédente :

Il s'agit d'approcher  $A$  par des techniques de rétro-analyse sur la seule base des effets produits.

La quantité d'information disponible peut être faible, les effets de  $A$  peuvent être furtifs, peu visibles lors de la période d'activation de  $A$  ou très intenses sur une très courte période (high frequency activity).

L'autorité supervisant  $A$  peut avoir un intérêt stratégique à ne pas divulguer le contenu de  $A$  ; il convient alors de procéder par cryptanalyse en particulier dans un contexte d'analyse virale ou de cyberdéfense.

La description de l'algorithme  $A$  s'effectue à différents niveaux :

- Le niveau « fonctionnel - complexité »
- Le niveau « stratégique - impact »
- Le niveau « évolution - adaptation »

### Niveau « fonctionnel - complexité » :

Ce niveau contient les caractéristiques techniques de l'algorithme ; on y retrouve :

Sa classe : Simulation numérique d'un système dynamique, modélisation par éléments finis, fonction résolvante de problèmes multi-contraints, analyseur d'événement par réseaux bayésiens, magma algorithmique de trading-HFT, algo-rédacteur, système cognitif, ....

Ses domaines de définition et de sortie : type de données fournies en entrée, type de données obtenues en sortie. Description détaillée du contexte d'utilisation de  $A$  et de la production de  $A$ .

Ses composantes : Liste complète des instructions algorithmiques composant  $A$  ; dans un langage de programmation classique (par le code source) ou via un langage formel  $L$  adapté au contexte.

Sa complexité : Longueur du code source représentant  $A$  dans le langage  $L$ , taille de l'espace mémoire nécessaire au fonctionnement de  $A$  sur une machine  $M$  donnée, temps d'exécution de  $A$  écrit dans un langage  $L$  et exécuté sur  $M$ , zone de stabilité de  $A$ , zone d'instabilité.

Interactions connues avec d'autres algorithmes,

Son domaine de performance : cela peut être l'échelle de temps associée à l'exécution de  $A$  ; un algorithme HFT travaillera sur une très courte période temporelle de l'ordre de la milliseconde, alors qu'un algorithme de simulation d'évolution de systèmes de galaxies pourra « tourner » durant une année ou plus sur un super computer (HPC calcul haute performance [4]).

Cela peut être également le volume de données que  $A$  prend en entrée : c'est le cas d'un algorithme  $A$  analysant un très grand nombre de tweets dans le but de dégager une tendance globale puis de proposer une prévision à partir de bigdatas.

La performance peut encore se situer sur la capacité de  $A$  à fournir des sorties proches de ce qu'un opérateur humain pourrait fournir sur la même question.

Il s'agit alors d'une performance d'imitation humaine ou de mimétisme (on pense aux robots téléphoniques de centre d'appels).

Les domaines de performance de A feront l'objet d'une évaluation fine s'appuyant sur des indicateurs numériques tangibles.

On construira des hauteurs de performance calculées sur la base de tests d'usage de A, si celui-ci est disponible. Dans le cas contraire (situation 2), on approchera ces hauteurs par la seule analyse des effets visibles de A sur l'environnement d'application E.

On pourra définir la hauteur de Turing de A notée  $HT(A)$  ; celle-ci mesurera la capacité de A à produire ou mimer des réponses humaines.

On évaluera  $HF(A)$ , la hauteur de furtivité de l'algorithme A, à partir de sa capacité à dissimuler son action sur E.

Ces hauteurs feront l'objet de définitions précises et formelles. Chaque caractéristique technique de A donnera lieu à une évaluation numérique par une hauteur ad-hoc.

On formera enfin une hauteur globale du niveau « fonctionnel-complexité » basée sur un modèle équationnel à définir, intégrant logiquement chaque hauteur-composante.

### **Niveau « stratégique-impact »**

On utilisera les hauteurs de compétences évaluées dans le niveau précédent pour décrire le potentiel stratégique de A et sa largeur d'impact sur l'environnement d'application E.

Là encore, il faudra construire un modèle équationnel pertinent et multi-paramètres donnant une image précise de la dimension stratégique liée à A.

Le niveau « stratégique-impact » devra évaluer la capacité disruptive de A au regard du magma algorithmique existant et agissant sur le même environnement E.

On mesurera tout d'abord le progrès induit par A en terme d'efficacité (vitesse, pertinence, stabilité, moyens de calcul mobilisés), on évaluera ensuite les effets collatéraux sur E émergeant d'une utilisation intensive de A. Cette évaluation est certainement la plus complexe à mettre en œuvre mais elle s'avère strictement nécessaire dans notre tentative d'accompagner la convergence et la future singularité.

### **Niveau « évolution-adaptation »**

On s'interrogera ici sur la capacité de notre algorithme A à évoluer, via adaptation, vers un environnement d'application E' distinct de l'environnement initial E.

On évaluera l'effort algorithmique à fournir pour que A devienne pertinent sur E' (alors qu'il l'est déjà sur E). On calculera ensuite le gain obtenu après action de A sur E' toujours au regard de l'existant algorithmique agissant sur E'.

Illustrons ce niveau avec deux cas concrets :

Un code viral A est détecté et analysé selon le niveau « fonctionnel-complexité ».

On cherche alors à modifier le contenu de A pour obtenir un code A' pertinent sur un nouvel environnement d'application E' (et sans charge virale pour E').

Un algorithme B réalisant des classements intelligents par compression est utilisé en génétique.

On évalue les modifications à apporter à B pour l'utiliser lors de classements musicaux, botaniques ou historiques, on peut ensuite dériver cet algorithme en un code efficace d'intelligence artificielle.

Certains algorithmes pourront être engendrés par des algorithmes parents; le niveau « évolution-adaptation » offrira alors l'arbre généalogique de A en incluant ses mutations et adaptations.

La construction d'une structure d'OEA constitue un premier pas fondateur dans notre préparation aux changements exponentiels et à une éventuelle singularité technologique.

Il convient de compléter cette création par des actions de sensibilisation et de formation à destination des jeunes générations (en particulier redonner le goût des sciences aux plus jeunes).

Les changements disruptifs et la convergence les impacteront de plein fouet, engendreront des instabilités inconnues jusqu'à présent et les contraindront à une constante remise en question de l'acquis intellectuel.

Il est temps d'agir aujourd'hui, avant que le code ne le fasse pour nous...

## Notes et renvois

[1] Singularité et technoprophéties- Thierry Berthier – alliance géostratégique.

<http://alliancegeostrategique.org/2012/12/04/singularites-et-techno-propheties/>

[2] Le cygne noir, la puissance de l'imprévisible – Nassim Nicholas Taleb – Les belles lettres.

[3] université de la singularité

<http://singularityu.org/>

[4] Supercomputer activity – HPC – Thierry Berthier – alliance géostratégique.

<http://alliancegeostrategique.org/2013/01/28/supercomputer-activity/>

## Illustrations

### Sentinelle SGRA1 Samsung



### Flash Crash Gaz HFT



### Flash Crash Or HFT



## Compte-rendu du match Angels / Red Sox Oct. 11 playoff game rédigé par un algo-rédacteur Stats Monkey

BOSTON — Things looked bleak for the Angels when they trailed by two runs in the ninth inning, but Los Angeles recovered thanks to a key single from Vladimir Guerrero to pull out a 7-6 victory over the Boston Red Sox at Fenway Park on Sunday. Guerrero drove in two Angels runners. He went 2-4 at the plate. “When it comes down to honoring Nick Adenhardt, and what happened in April in Anaheim, yes, it probably was the biggest hit (of my career),” Guerrero said. “Because I’m dedicating that to a former teammate, a guy that passed away.” Guerrero has been good at the plate all season, especially in day games. During day games Guerrero has a .794 OPS. He has hit five home runs and driven in 13 runners in 26 games in day games. After Chone Figgins walked, Bobby Abreu doubled and Torii Hunter was intentionally walked, the Angels were leading by one when Guerrero came to the plate against Jonathan Papelbon with two outs and the bases loaded in the ninth inning. He singled scoring Abreu from second and Figgins from third, which gave Angels the lead for good.

The Angels clinched the AL Division Series 3-0. Angels starter Scott Kazmir struggled, allowing five runs in six innings, but the bullpen allowed only one runs and the offense banged out 11 hits to pick up the slack and secure the victory for the Angels. J.D. Drew drove in two Red Sox runners. He went 1-4 at the plate. Drew homered in the fourth inning scoring Mike Lowell. “That felt like a big swing at the time,” said Drew. “I stayed inside the ball and put a good swing on it. I was definitely going to be ready to battle again tomorrow, but it didn’t work out.” Drew has been excellent at the plate all season, especially in day games. During day games Drew has a .914 OPS. He has hit five home runs and driven in 17 runners in 36 games in day games. Papelbon blew the game for Boston with a blown save. Papelbon allowed three runs on four hits in one inning. Reliever Darren Oliver got the win for Los Angeles. He allowed no runs over one-third of an inning. The Los Angeles lefty struck out none, walked none and surrendered no hits. Los Angeles closer Brian Fuentes got the final three outs to record the save. Juan Rivera and Kendry Morales helped lead the Angels. They combined for three hits, three RBIs and one run scored. Four relief pitchers finished off the game for Los Angeles. Jason Bulger faced four batters in relief out of the bullpen, while Kevin Jepsen managed to record two outs to aid the victory.