



# intelligence artificielle & espace public

automne 2024

Cycle de conférences publiques dans le cadre du cours de Master en Humanités numériques et de Master en Sciences sociales « Théorie de la culture et des médias »

21 octobre 2024 | 16h15-18h00 | Internef 272

**La dialectique de la raison algorithmique :  
Pour une théorie critique de l'intelligence  
artificielle**

**Maxime Ouellet**, UQAM, Montréal

11 novembre 2024 | 16h15-18h00 | Internef 272

**L'intelligence artificielle comme projet de  
normativisation : Vers de nouvelles formes de  
pouvoir et de subjectivité**

**Katia Schwerzmann**, Thyssen Fellow au  
Kulturwissenschaftliches Institut à Essen, Allemagne

13 novembre 2024 | 12h15-14h00 | Géopolis 2230

**Esprit et information**

**Mark Hunyadi**, Université catholique de Louvain, Bruxelles

18 novembre 2024 | 16h15-18h00 | Internef 272

**Les secrets de la transparence.**

**L'exemple du moteur de recherche**

**Thomas Berns**, Université libre de Bruxelles

25 novembre 2024 | 16h15-18h00 | Internef 272

**3038 (50.97) Un art de machine ?**

Antòn de Macedo, Collectif Aristide &  
**Paul Hegi**, photographe (dipl ECAL)

2 décembre 2024 | 16h15-18h00 | Internef 272

**Apprendre à interdire. Gouvernance par le  
prompt et la censure des IA génératives**

**Ksenia Ermoshina**, Université de Toronto & CNRS

**Université de Lausanne  
Bâtiments Internef et Géopolis**

Organisation: Olivier Voirol, Olivier Glassey et  
Valentine Girardier

Pour toute information : [olivier.voirol@unil.ch](mailto:olivier.voirol@unil.ch)



## intelligence artificielle & espace public

cycle de conférences publiques - automne 2024

### Présentation du cycle

La question de l'intelligence artificielle est au cœur des préoccupations actuelles au sein de l'espace public. Son application au sein des secteurs d'activité les plus divers fait (re)surgir le spectre de la machine capillaire douée d'une calculatoire propre, nourrie de capacités hors pair d'analyse, d'apprentissage et de décision. Une panoplie d'acteurs y voient toutefois une promesse d'efficacité portée par la rationalité de l'automation, propre à revigorer le développement du capitalisme. Les sociétés modernes démocratiques se sont constituées depuis le 18<sup>e</sup> siècle sur la base d'une aspiration à la raison et à l'autoréflexion. Celle-ci se fonde sur des aptitudes collectives à moduler les choix et les pratiques. Le spectre d'une « société automate » (Stiegler) que fait planer l'IA, régie par des robots autonomisés sur lesquels les êtres sociaux perdent toute capacité d'action et de décision, est aux antipodes de l'idée d'espace public sur laquelle reposent les sociétés modernes. L'espace public est actuellement sous tension face à ce spectre de l'automation. Une des questions qui se posent à l'heure actuelle est de savoir comment l'espace public peut « reprendre la main » sur l'IA, en la soumettant à une culture critique, aux régulations éthico-juridiques et à l'agir politique.

### Maxime Ouellet : La dialectique de la raison algorithmique : Pour une théorie critique de l'intelligence artificielle

À partir de la Théorie critique développée initialement par l'Ecole de Francfort, cette présentation a comme ambition de saisir les fondements sociohistoriques des catégories centrales qui sont au fondement de l'intelligence artificielle : la communication, la commande, le contrôle et l'information. Nous situons tout d'abord les développements de l'intelligence artificielle dans le sillage des premiers travaux issus de la cybernétique dans le cadre de ce qui était qualifié à l'époque de capitalisme monopoliste d'Etat. Nous poserons ensuite la question visant à savoir si les mutations contemporaines du capitalisme rendues possibles grâce aux avancées dans le domaine de l'intelligence artificielle tel que le machine learning et le deep learning remettent en question la possibilité de développer une théorie critique de la société.

### Mark Hunyadi - Esprit et information

Intelligence artificielle et numérique relèvent du même paradigme scientifique : la cybernétique. Et de fait, dans le monde numérique, chaque individu est appelé à devenir, malgré lui, un opérateur cybernétique. Quel impact cela a-t-il sur la vie de l'esprit, qu'il soit individuel ou collectif ?

### Anton de Macedo & Paul Hegi: 3038 (50.97) un art de machine ?

Le duo d'artistes Paul Hegi & Anton de Macedo a publié cette année un livre d'art entièrement généré par l'intelligence artificielle (IA). Ce projet qui a nécessité deux ans de travail a dû relever de nombreux défis tels que développer un langage probant avec la machine, créer un corpus de plus de 1 '000 images avec éléments graphiques au service du narratif, mettre au point un nouveau flux de travail et d'imprimer sous forme d'ouvrage des images dédiées à la luminosité de l'écran. Ce livre soulève de nombreuses questions dans son contenu, tout d'abord, en posant la question de notre rapport à la technologie et de l'extinction du vivant, et ensuite dans son procédé reliant IA et création artistique. Le duo d'artiste a créé ce livre en premier lieu comme un outil-témoignage d'une époque, qui leur permet de continuer à développer l'indispensable débat autour de l'IA. Langage avec la machine, créativité, intérêt des images produites par ordinateurs, droit d'auteur, les récurrences de sémiotique visuelle, etc. C'est pourquoi, à la suite du livre, le duo d'artistes a créé plusieurs expositions tirées de ce dernier, afin d'expérimenter différentes façons d'exploiter ce corpus d'images.

### Thomas Berns: Les secrets de la transparence : l'exemple du moteur de recherche

Il y a des liens paradoxaux entre transparence et secret apparaissent à partir de l'exemple du moteur de recherche et, plus largement, de l'émergence de nouvelles formes de normativité, entre autres de nature technique, qui sont analysées depuis leur différence par rapport à la normativité juridique. Ce faisant, c'est aussi le danger de la dilution du principe de publicité dans l'idée de la transparence qui est rendue visible.

### Katia Schwerzmann: L'intelligence artificielle comme projet de normativisation : vers de nouvelles formes de pouvoir et de subjectivité

Cette présentation a pour objet la façon dont l'intelligence artificielle dans sa forme actuelle, que je conçois comme un projet de normativisation, participe à la constitution de nouvelles formes de pouvoir et de subjectivité. Il s'agira dans un premier temps de comprendre les implications de la transition d'un paradigme de programmation informatique basé sur la règle vers celui du machine learning, ou apprentissage automatisé, fondé sur l'exemple. Bien que ces deux paradigmes coexistent dans la pratique, cette transition a des implications épistémiques, éthiques et politiques importantes : règles et exemples régulent en effet les conduites humaines de manière différente. Contrairement à l'autorité explicite et prescriptive imposée par les règles, les modèles de machine learning produisent une autorité difficile à contester car basée sur des normes implicites en constante mutation. J'examinerai dans un deuxième temps comment les « large language models » tels que ChatGPT ou Claude produisent une forme de subjectivité spécifique – le « je » prononcé par ces modèles – via des opérations normatives d'« alignement » dont le but est de résoudre d'un même coup des problèmes techniques et éthiques. Ce « je » qui se positionne en porteur de savoir fait la performance d'une conception spécifique du savoir et de la subjectivité qu'il s'agira d'analyser de façon critique.

### Ksenia Ermoshina: Apprendre à interdire : gouvernance par le prompt et la censure des IA génératives

La démocratisation d'accès aux outils de génération d'images comme Dall-E, Midjourney ou Stable Diffusion, a suscité le besoin de régulation et d'encadrement des usages de ces IAs. Nous appelons cela "gouvernance par le prompt" qui consiste à interdire les usages de certains mots-clés dans les prompts textuels. Ces prompts interdits évoquent le plus souvent des registres lexicaux qui renvoient à la violence, nudité, sexualité, sexisme, racisme. Or, dans d'autres cas, les interdictions peuvent également concerner des personnalités politiques, des noms de pays ou de régions, des événements historiques ou même des identités sexuelles. Cette intervention est basée sur une enquête de terrain comparative autour des trois IA génératives à destination du grand public : le projet chinois ERNIE-ViLG, le projet russe Randinsky XXL et le projet américain DALL-E. Elle explore les impacts de facteurs culturels et politiques ainsi que des modes de production de ces IA et leurs licences sur les formes de censure et de « gouvernance par le prompt ». L'enquête explore également la propagation de pratiques de ruse et bricolage collectives développées par les utilisateurs face à la censure. Tout d'abord, pour mettre la lumière sur l'existence même des listes noires de prompts, mener des enquêtes collectives de "reverse engineering" afin de rétablir ces listes et les partager avec les autres. - Ensuite, pour trouver des façons de contourner ces interdictions, surtout par les artistes visuels.