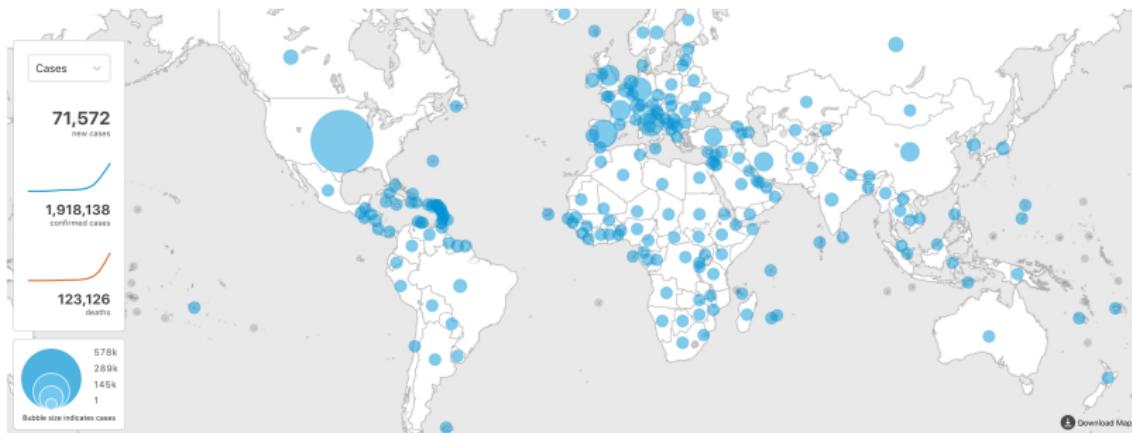


Un panorama des recherches contre le Covid-19 dans le domaine des sciences et technologies de l'information

Jamal Atif, Stéphane Canu et Olivier Cappé
github.com/StéphaneCanu/covid_AI_resources

« Journée d'étude : Covid 19 »



16 avril 2020

Contre le covid : recherches à tous les étages

Niveau	Domaines	Exemples	
Population	 $I \xrightarrow{\gamma} S$ $I \xrightarrow{\alpha} R$	Épidémiologie, information	Modèles, RI, tracer les contacts
Hôpital		Gestion de l'hôpital et des patients	SI, gestion des lits, essais cliniques
Patient		Aide au diagnostic	Analyse d'images de scan,
Molécule		Génomique, vaccin médicaments	Découverte de molécules

Road map

1 Introduction

- Structuration nationale et internationale
- La production et gestion des données et d'informations

2 Observation et Modélisation de la pandémie

3 A l'échelle des hopitaux

4 Au niveau du patient (aide au diagnostic)

5 Au niveau des molécules

- A la recherche de traitements
- Génomique épidémiologique

6 Partage des connaissances (infodémiologie ?)

- Gestion des informations
- Préparer la sortie du confinement

Structuration nationale

- Le comité analyse, recherche et expertise (CARE) centralise¹ : les thérapies, les diagnostics et, sur une autre échelle de temps, les futurs vaccins. Apporter un éclairage rapide sur les innovations scientifiques, thérapeutiques et technologiques
- INSERM : Reacting : coordonner la réponse de la recherche française²
- ANR : Appel flash Covid-19 (une dizaine de projets dans le domaine)
- Mission Inria Covid-19³
- Care@CNRS⁴
- ... et aussi : la task force Coronavirus de l'institut pasteur, l'appel à projets de la DGA, MODCOV19 au CNRS, l'Institut Français de Bioinformatique...

¹ <https://fr.surveymonkey.com/r/CARE-Covid19>

² <https://reacting.inserm.fr/>

³ www.cnrs.fr/fr/cnrsinfo/modcov19-la-modelisation-pour-mieux-lutter-contre-la-pandemie

⁴ <http://www.cnrs.fr/en/node/4620>

Initiatives internationales

- Workshop Ellis against covid19⁵
- Workshop Stanford⁶
- Kaggle (the COVID-19 Open Research Dataset Challenge)⁷
- L'europe⁸
- Mapping the Landscape of Artificial Intelligence Applications against COVID-19⁹ (25 mars 2020)



⁵ <https://ellis.eu/covid-19#livestreaming>

⁶ <https://hai.stanford.edu/events/covid-19-and-ai-virtual-conference/agenda>

⁷ <https://www.kaggle.com/covid-19-contributions>

⁸ <https://www.coe.int/fr/web/artificial-intelligence/ai-and-control-of-covid-19-coronavirus>

⁹ <https://arxiv.org/abs/2003.11336>

Les tableaux de bord liés à l'épidémie

- Monde (OMS)

- ▶ Worldometers¹⁰
- ▶ John Hopkins dashboard¹¹

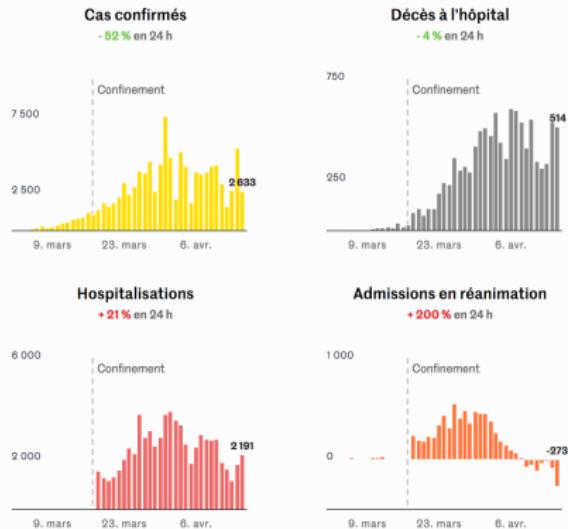
- France

- ▶ Santé Publique France¹²
- ▶ COVID-19 en France¹³
- ▶ Le Monde¹⁴
- ▶ tiny.cc/covidfr...

- Suisse¹⁵

- UK¹⁶

- US¹⁷...



10 <https://www.worldometers.info/coronavirus/>

11 <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>

12 <https://www.santepubliquefrance.fr>

13 <https://dashboard.covid19.data.gouv.fr/>

14 https://www.lemonde.fr/les-decodeurs/article/2020/02/27/en-carte-visualisez-la-propagation-mondiale-de-l-epidemie-de-coronavirus_6031092_4355770.html

15 <https://covid-19-schweiz.bagapps.ch/fr-1.html>

16 <https://imperialcollegelondon.github.io/covid19estimates>

17 <https://covid19.healthdata.org/france>

Les données liées à l'épidémie

- France : opencovid19-fr/data¹⁸
- Monde : (kaggle)
 - ▶ brutes : github.com/beoutbreakprepared/nCoV2019¹⁹
 - ▶ consolidées : covid.ourworldindata.org²⁰
 - ▶ the COVID-19 Open Research Dataset: 51,000 scholarly articles
- autres sources
 - ▶ décès par jour et par tranche d'âge et genre²¹
 - ▶ social contact matrices in 152 countries using contact surveys and demographic data. Kiesha Prem, Alex R. Cook (2017)²²
 - ▶ mobilité géographiques, démographie (dc-covid.site.ined.fr)

	date	granularité	maille_code	maille_nom	cas_confirmés	déces	déces_ehpad	réanimation	hospitalisés	guéris	dépistés	source_nom	sou
5917	2020-04-09	region	REG-75	Nouvelle-Aquitaine	0.0	173.0	NaN	241.0	792.0	789.0	NaN	OpenCOVID19-fr	
5918	2020-04-09	region	REG-76	Occitanie	0.0	203.0	NaN	328.0	992.0	1081.0	NaN	OpenCOVID19-fr	
5919	2020-04-09	region	REG-84	Auvergne-Rhône-Alpes	0.0	685.0	NaN	720.0	2962.0	2508.0	NaN	OpenCOVID19-fr	
5920	2020-04-09	region	REG-93	Provence-Alpes-Côte d'Azur	0.0	286.0	NaN	426.0	1748.0	1641.0	NaN	OpenCOVID19-fr	
5921	2020-04-09	region	REG-94	Corse	0.0	34.0	NaN	19.0	88.0	112.0	NaN	OpenCOVID19-fr	

18

<https://raw.githubusercontent.com/opencovid19-fr/data/master/dist/chiffres-cles.csv>

19

https://raw.githubusercontent.com/beoutbreakprepared/nCoV2019/master/latest_data/latestdata.csv

20

https://covid.ourworldindata.org/data/ecdc/full_data.csv

21

<https://dc-covid.site.ined.fr/fr/donnees/>

22

<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005697>

La recherche épidémiologique et la modélisation

- La référence : Neil Ferguson, imperial college²³
 - ▶ semi-mechanistic Bayesian hierarchical model (leur github)²⁴
- aux US :
 - ▶ SEIR discret à compartiment (John Hopkins) ²⁵
 - ▶ Delphi prediction team (epiforecast soft, CMU)²⁶
- en france :
 - ▶ COVIDSIM, Groupe de modélisation, MIVEGEC, CNRS, IRD²⁷
 - ▶ Modèle français par région, Pacal Crepey²⁸
 - ▶ ANR covid MPCUII : modèles et prévisions pour COVID-19 avec des individus infectieux non déclarés, Pierre Magal²⁹
 - ▶ ANR covid NoCOV Prévisions au court et moyen terme de la diffusion de COVID-19, Chiara Poletto

23 <https://imperialcollegelondon.github.io/covid19estimates>

24 <https://github.com/ImperialCollegeLondon/covid19model>

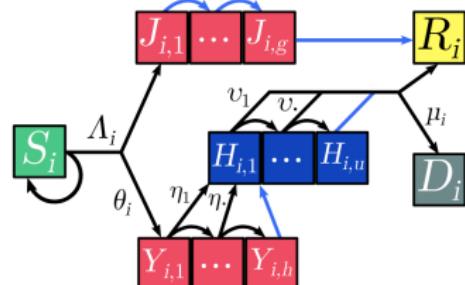
25 <https://www.nature.com/articles/s41598-019-38665-w>

26 <https://delphi.cmu.edu/>

27 <http://bioinfo-shiny.ird.fr:3838/COVIDSIM/>

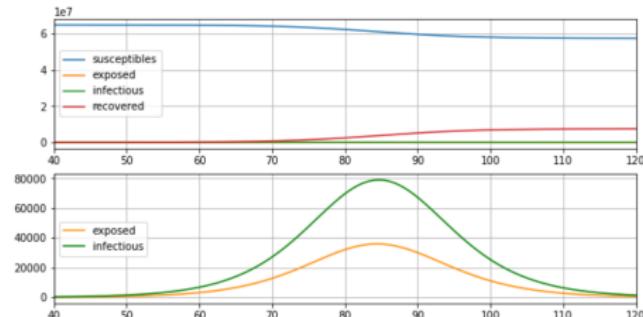
28 <https://github.com/PascalCrepey/HospCoV>

29 <https://www.math.u-bordeaux.fr/~pmagal100p/LMSW-coronai-MPDIBiology.pdf>



La recherche épidémiologique et la modélisation

- Modèles alternatifs : Modified Auto-encoder³⁰
- Intégration des contraintes géographiques à grain fin,
 - ▶ Projet ANR flash Covid-19 : EVALCOVID-19 Modélisation mathématique et numérique intégrant des données de téléphonie mobile en temps réel pour évaluer les interventions contre la pandémie de COVID-19 en France (Vittoria Colizza)
- Estimation des paramètres et assimilation de données
 - ▶ travail sur les modèles SIR (Marc Lavielle)³¹
 - ▶ Estimation du taux de transmission en fonction du temps
- Gestion des erreurs
- Recherche de la saisonnalité
- ...



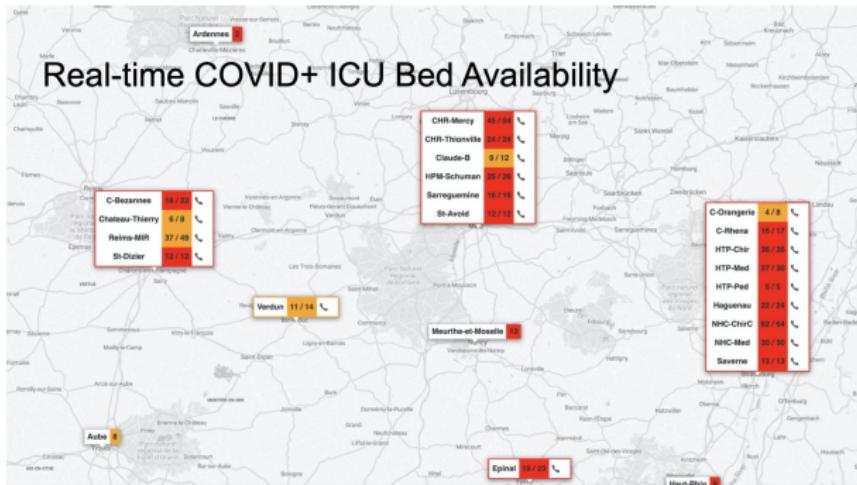
³⁰

Z. Hu et al.. Artificial intelligence forecasting of COVID-19 in China. arXiv preprint arXiv:2002.07112, 2020

³¹

<http://webpopix.org/covidix19.html>

A l'échelle des hopitaux



- ICUBAM API d'information en temps réel sur la disponibilité des lits³² (open-source GitHub, INRIA, X, CHRU Nancy & U. de Lorraine)
- Gestion des appels d'urgence (réseau de pétri dynamique et optimisation), V. Bœuf, X. Allamigeon et S. Gaubert
- Aide à l'analyse de données cliniques Covid-19

³² <https://icubam.github.io/>

L'aide au diagnostic

- Analyse d'image du poumon

- ▶ des images³³ et le réseau covid imagerie SFR
- ▶ modified inception architecture³⁴
- ▶ UNet++ architecture³⁵
- ▶ infervision, Delft Imaging (CAD4COVID)

- Analyse de la respiration (des mouvements)

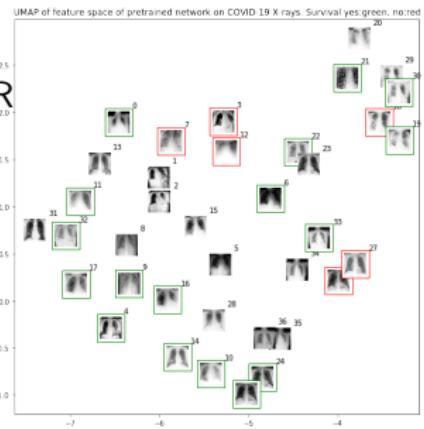
- ▶ Kinect pour identifier des schémas respi.³⁶

- Analyse de la toux (du son)

- ▶ Coughvid (EPFL, LauzHack)³⁷
- ▶ AI4COVID-19 App (pas encore disponible)

- Estimation des risques

- ▶ La sévérité avec XGBoost³⁸ , durée de séjour avec Unet³⁹
- ▶ Proposition d'un index de vulnérabilité avec XGboost⁴⁰



33 <https://github.com/ieee8023/covid-chestxray-dataset>

34 S. Wang et al., A deep learning algorithm using CT images to screen for corona virus disease

35 J. Chen et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography...

36 Y. Wang et al. Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19...

37 <https://coughvid.epfl.ch/>

38 L. Yan et al. Prediction of criticality in patients with severe COVID-19 infection using three clinical features: a ML-based prognostic model...

39 X. Qi et al. ML-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection

40 D.DeCappi et al. Building a COVID-19 Vulnerability Index

L'aide à l'échelle des molécules

- Prédiction de la structure des protéines du virus
 - ▶ deep learning : Linearfold (2019)⁴¹, AlphaFold (2019)⁴²
 - ▶ Folding@Home de l'université Stanford (simulate protein in action)⁴³
- ANR Covid 19 PullCoVapart : Combine la biochimie des protéines avec des ontologies pour proposer des options thérapeutiques (I. Imbert)
- Amélioration des tests (dépistage et sérologie)⁴⁴
- Réutilisation de médicaments
 - ▶ Utiliser des graphes de connaissances biomédicales⁴⁵
 - ▶ The Joint European Disruptive Initiative (JEDI) : rapidly identify molecules/peptides/proteins that can help (screening a billion molecules against SARS-CoV-2 in 4 weeks)⁴⁶



41 LinearFold: linear-time approximate RNA folding by 5'-to-3' dynamic programming and beam search

42 A. Senior et al. Improved protein structure prediction using potentials from deep learning, *Nature*, 15 jan 2020

43 <https://foldingathome.org/covid19/>

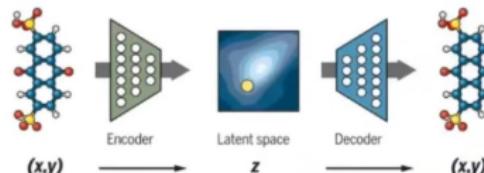
44 H. Metsky et al. CRISPR-based surveillance for COVID-19 using genomically-comprehensive machine learning design

45 M. Segler et al. Planning chemical syntheses with deep neural networks and symbolic AI

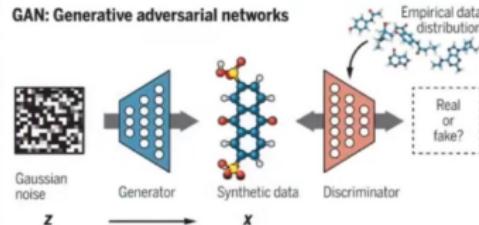
46 <https://www.covid19.jedi.group/>

Découverte de nouveaux médicaments (drug discovery)

VAE: Variational autoencoders

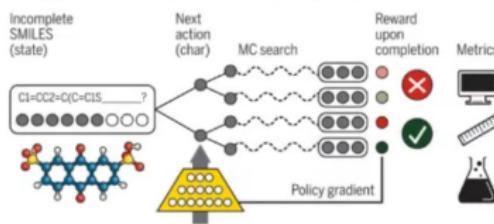


GAN: Generative adversarial networks



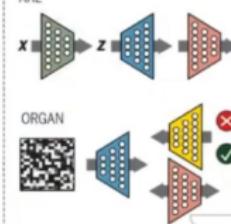
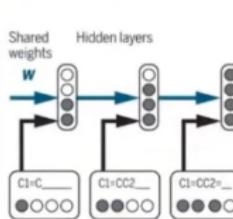
RL: Reinforcement learning

Policy gradient with Monte Carlo tree search (MCTS)



RNN: Recurrent neural network

Hybrid approaches



- Competition Kaggle⁴⁷
- Open and Collaborative (OpenAI)⁴⁸
- Entre autres, Insilico propose 6 molécules le 6/02/20 (2 synthétisés) ⁴⁹

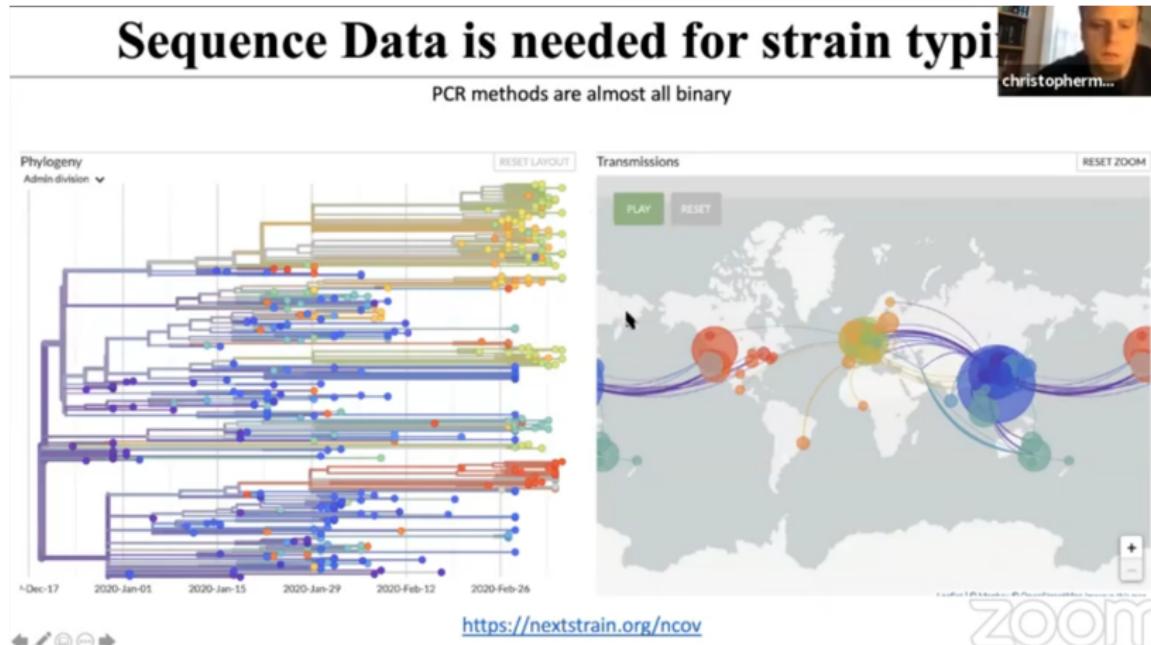
⁴⁷ <https://www.kaggle.com/maria17/covid-19-explore-drugs-being-developed>

⁴⁸ <http://www.montreal.ai/opencovid19.pdf>

⁴⁹ <https://insilico.com/ncov-sprint>

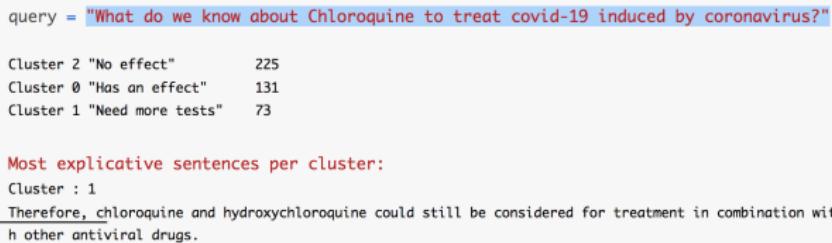
Génomique épidémiologique

- Partage des genomes du virus via open access databases⁵⁰
- Visualiser l'évolution du génome du virus



La gestion des informations (langage naturel)

- Analyse bibliographique des articles scientifiques sur la pandémie
 - ▶ découverte de tendance : COVID-19 Primer⁵¹
 - ▶ ANR Covid living-NMA-covid, Isabelle Boutron⁵²
- Kaggle challenge : NLP to answer key questions from the literature
 - ▶ données : Allen's COVID-19 Open Research Dataset (CORD-19)⁵³
 - ▶ premiers résultats : Kaggle Literature Review⁵⁴
 - ▶ exemple de contribution : BERT + MeSH ontology⁵⁵
 - ▶ un autre exemple : topic modeling, finding related articles⁵⁶
 - ▶ encore un autre : analyse d'opinions (Atos/Bull Fr)⁵⁷



51 <https://covid19primer.com/>

52 <https://covid-nma.com>

53 <https://pages.semanticscholar.org/coronavirus-research>

54 <https://www.kaggle.com/covid-19-contributions>

55 <https://www.kaggle.com/sandyvarma/covid-19-bert-mesh-enabled-knowledge-graph>

56 <https://discovid.ai/>

57 <https://github.com/MrMimic/covid-19-kaggle>

NLP (suite)

- Suivit des réseaux sociaux

- ▶ La société BlueDot détecte le 31/12/19
 - ★ HealthMap et Taïwan aussi...
- ▶ Projet ANR flash Covid-19 : DigEpi, analyse des réseaux sociaux pour anticiper le niveau de transmission dans les agglomérations (B. Roche)
- ▶ Comment les informations vraies ou fausses se propagent ?
 - ★ Projet ANR flash Covid-19 : TRACTRUST surveiller la confiance et la méfiance: analyser les médias sociaux [...] (L. Atlanti-Duault)
 - ★ Analyse multi linguistique⁵⁸
 - ★ Modélisation et estimation du taux de transmission R_0 ⁵⁹



- Agents conversationnels

- ▶ versions « officielles » : OMS, gouvernement, Pasteur, WhatsApp, facebook Google⁶⁰
- ▶ covid-bot par la société Clevy (github open source)⁶¹

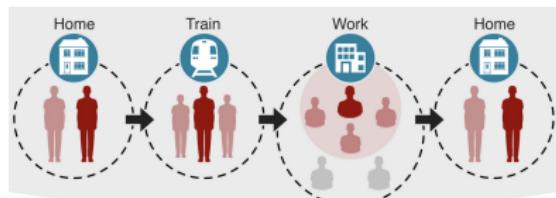
⁵⁸ C. Lopez et al. Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset

⁵⁹ M. Cinelli et al. The COVID-19 Social Media Infodemic

⁶⁰ <https://github.com/GoogleCloudPlatform/covid19-rapid-response-demo>

⁶¹ <https://www.covidbot.fr/>

Préparer la sortie du confinement



- Quoi faire ? (je ne sais pas)
 - ▶ Tracer les contacts⁶²
 - ★ Idées de Y. Bengio sur le sujet⁶³
 - ★ En france : Expected impact of lockdown in Île-de-France...⁶⁴
 - ▶ Modéliser les sénarii possibles (aide à la décision)

- Comment faire ? Privacy-by-Design Solution

- ▶ 10 avril : Apple and Google joint Bluetooth and cryptography specifications for privacy-preserving proximity tracing⁶⁵
- ▶ Projet associatif PEPP-PT (Privacy-Preserving Proximity Tracing)⁶⁶
 - ★ The DP-3T projet : Decentralized Privacy-Preserving Proximity Tracing (github)⁶⁷
- ▶ Safe Path (MIT media lab, github)⁶⁸

⁶² L. Ferretti et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing

⁶³ <https://yoshuabengio.org/fr/2020/03/25/depistage-pair-a-pair-de-la-covid-19-base-sur-lia/>

⁶⁴ https://www.epicx-lab.com/uploads/9/6/9/4/9694133/inserm-covid-19_report_lockdown_idf-20200412.pdf

⁶⁵ <https://www.apple.com/covid19/contacttracing/>

⁶⁶ <https://www.pepp-pt.org/>

⁶⁷ <https://github.com/DP-3T/documents>

⁶⁸ <https://github.com/tripleblindmarket/covid-safe-paths>

Conclusions

- La recherche immédiate :
 - ▶ une mobilisation forte du monde de la recherche
 - ▶ rien n'est parfait... il reste sans doute quelques opportunistes
- Pour une recherche ouverte
 - ▶ un moment de vérité (par exemple en *drug discovery*)
 - ▶ des données réelles à l'échelle du globe (génome, images...)
- La recherche pour dans 6 mois (2nd vague cet automne ?)
 - ▶ pluridisciplinaire, ouvert,

Continuer à accélérer sur certains sujets :

Population NLP (RI et chatbot)

big data et apprentissage statistique

Patient aide au traitement des données cliniques (images, diagnostic...)

Modécule *drug discovery*