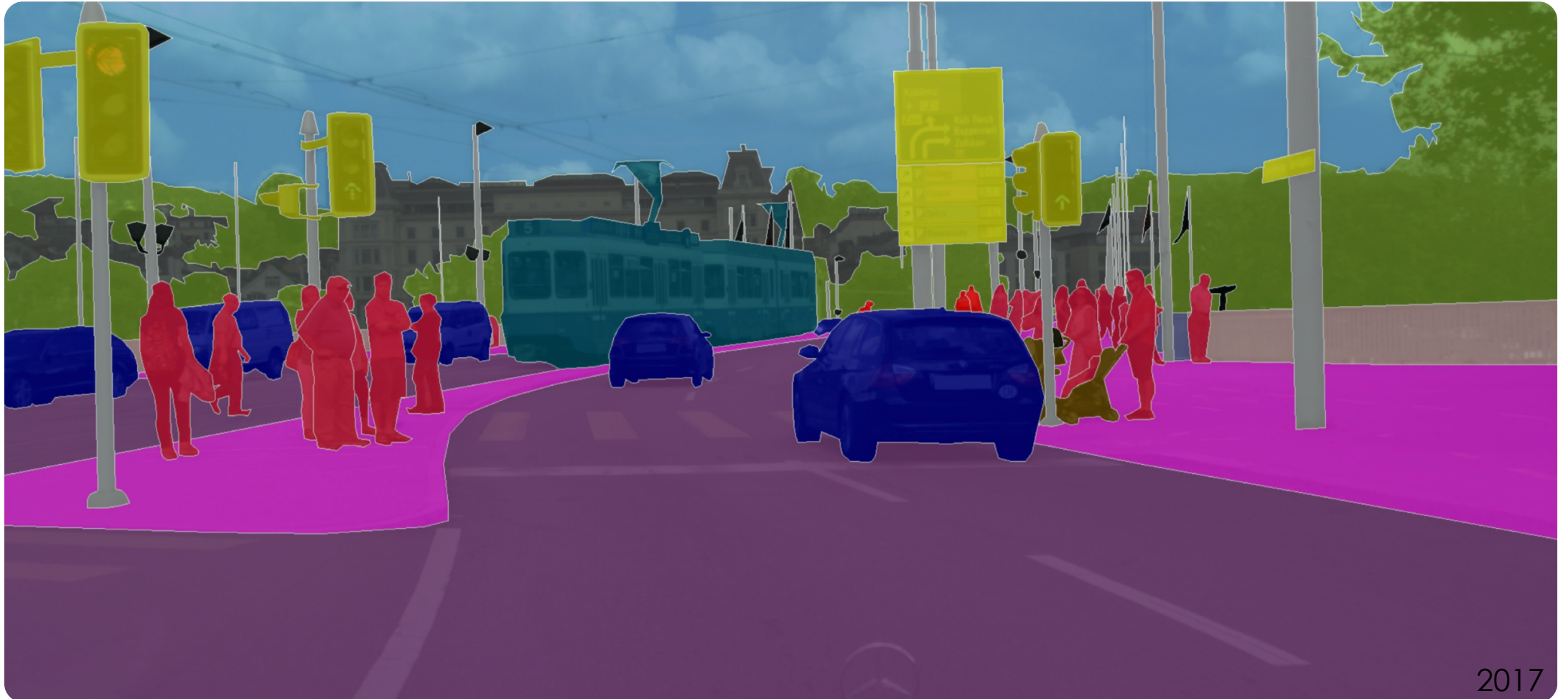


Apprentissage automatique et raisonnement causal

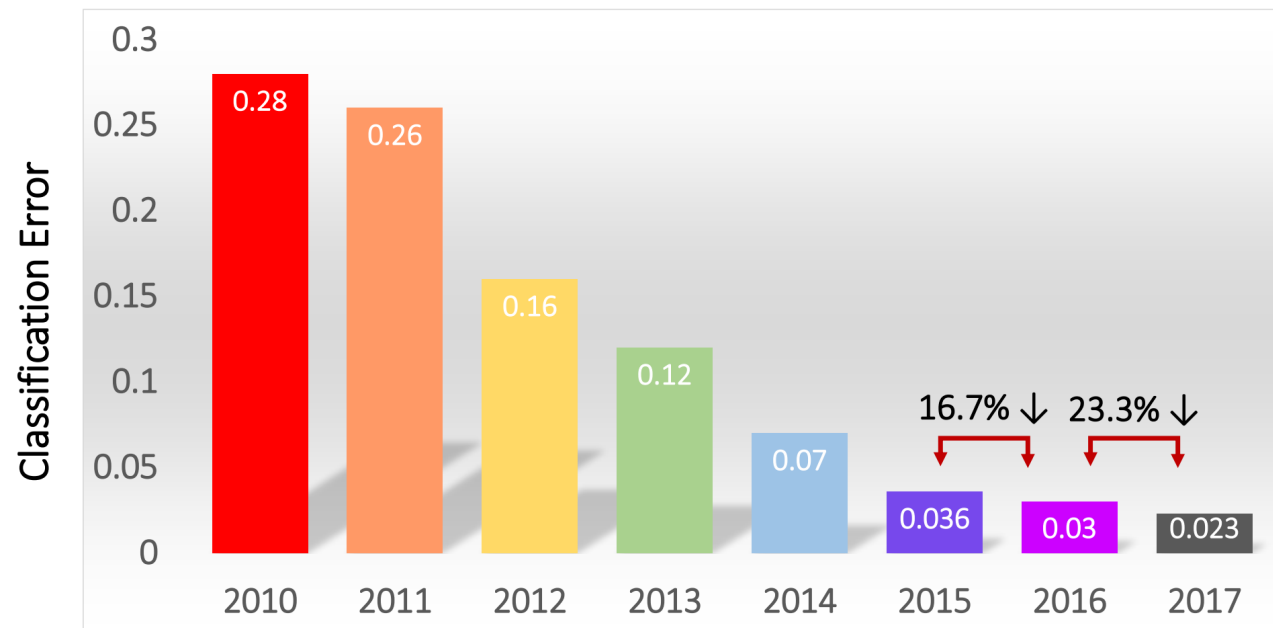
Au-delà des corrélations

L'IA a récemment fait des progrès époustouflants

De la perception, nous sommes passés au raisonnement



Les modèles sont devenus experts dans certaines tâches visuelles



Source: http://image-net.org/challenges/talks_2017/ILSVRC2017_overview.pdf

La compétition ImageNet originale a été discontinuée



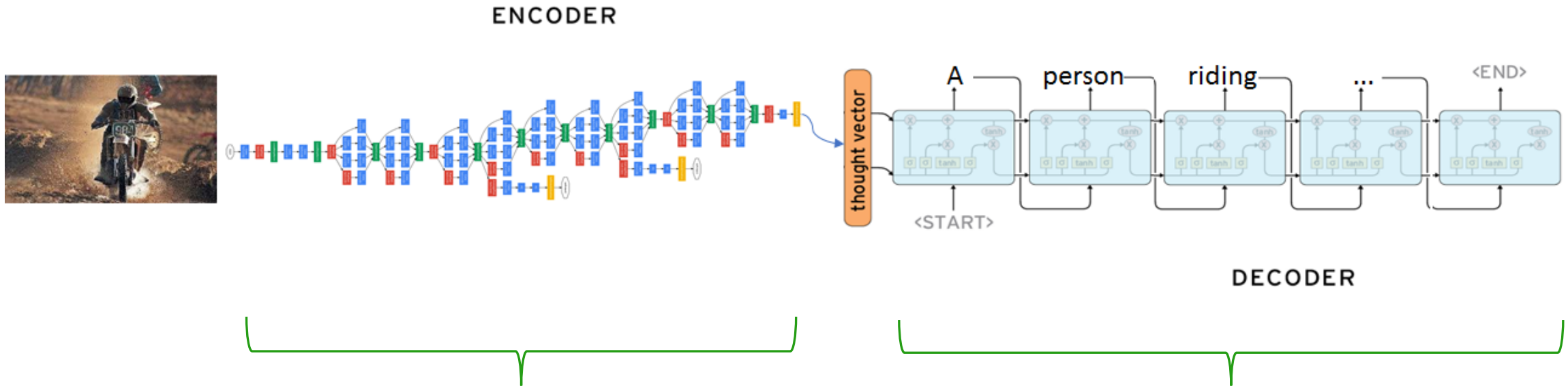
Modélisation de distributions en très haute dimension



Source: <https://arxiv.org/abs/1906.00446> (VQ-VAE 2)

Les réseaux de neurones profonds

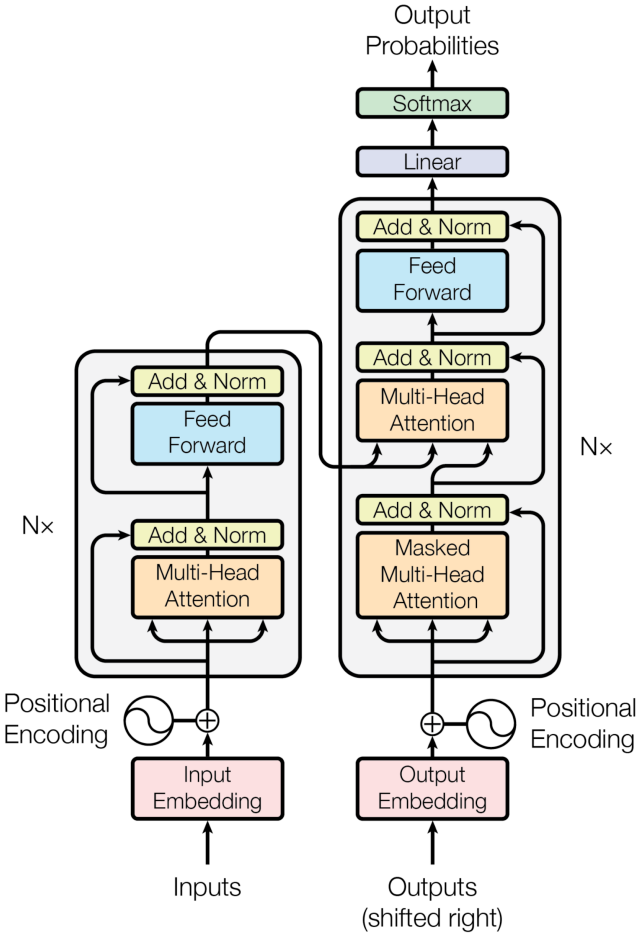
ont joué un rôle majeur dans ces avancées



**Convolutional Neural Network
(e.g. Inception)**

**Recurrent Neural Network
(e.g. LSTM)**

Les modèles *Transformer* sont à l'origine de plusieurs avancées récentes



Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

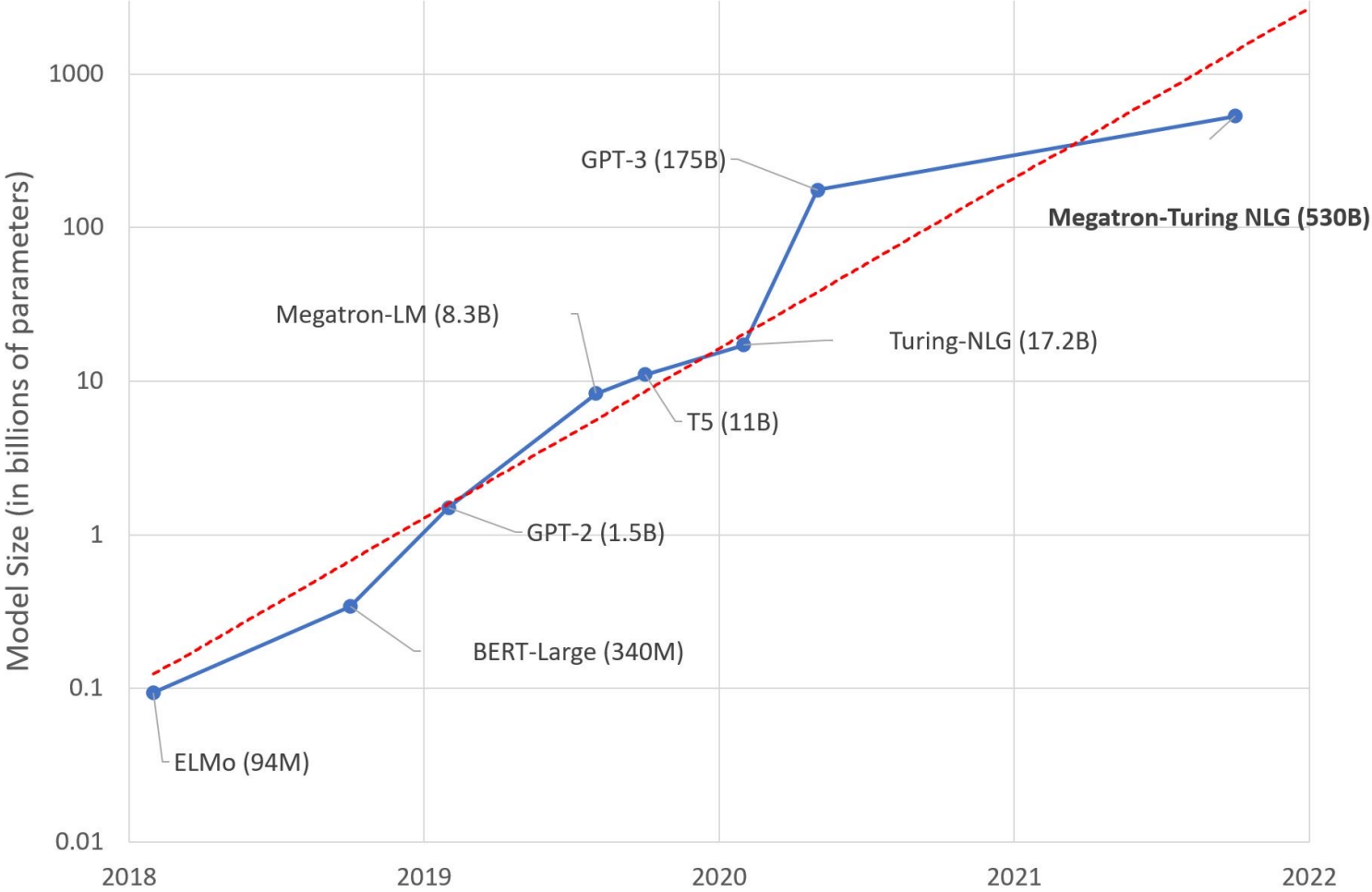
Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaizer@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

La puissance de ces modèles ne cesse d'augmenter

Des habiletés de raisonnement surprenantes émergent



On peut maintenant générer du texte réaliste

Québec, le 27 mars 2023 Le maire de la ville de Québec, Bruno Marchand, a dévoilé aujourd'hui un ambitieux projet de mobilité durable basé sur la téléportation. Ce projet, qui pourrait être mis en place d'ici 2025, vise à réduire considérablement les émissions de gaz à effet de serre et à améliorer la qualité de vie des citoyens.

Selon le maire Marchand, la téléportation représente l'avenir de la mobilité urbaine. Ce moyen de transport innovant permettrait de réduire la congestion routière, de diminuer la dépendance aux énergies fossiles et de favoriser un mode de vie plus écologique.

Le projet prévoit la construction de stations de téléportation réparties à travers la ville, offrant un accès facile et rapide aux usagers. Les déplacements se feraient instantanément, réduisant ainsi les temps de trajet et les émissions polluantes.

Les investissements nécessaires au développement de cette technologie seront répartis entre la ville de Québec, le gouvernement provincial et des partenaires privés. Le maire Marchand espère que ce projet novateur permettra de positionner la ville de Québec comme un leader mondial en matière de mobilité durable et d'innovation technologique.

Généré avec GPT-4

... traduire du texte en images

Input

An astronaut riding a horse in photorealistic style.



... générer du code

GitHub Copilot – Open AI Codex



```
10 def send_tweet_with_image(message, image):
11     """Send a tweet with an image attached"""
12     # Twitter authentication
13     auth = tweepy.OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET)
14     auth.set_access_token(ACCESS_TOKEN, ACCESS_TOKEN_SECRET)
15     api = tweepy.API(auth)
16
17     # Send the tweet with the image
18     api.update_with_media(image, status=message)
19
20
21
22
23 def fetch_tweets_from_user(username):
24     # authentication
25     auth = tweepy.OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET)
26     auth.set_access_token(ACCESS_TOKEN, ACCESS_TOKEN_SECRET)
27     api = tweepy.API(auth)
28
29     # fetch tweets
30     tweets = api.user_timeline(username, count=20, include_rts=False)
31     return tweets
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```



BigCode

Initiative **ouverte et responsable** pour l'élaboration de modèles de génération de code

Collaboration ServiceNow & HuggingFace, avec une communauté de **400 participants** provenant de **30 pays**.



Hugging Face + **servicenow**

26 sept., 2022:
Lancement officiel

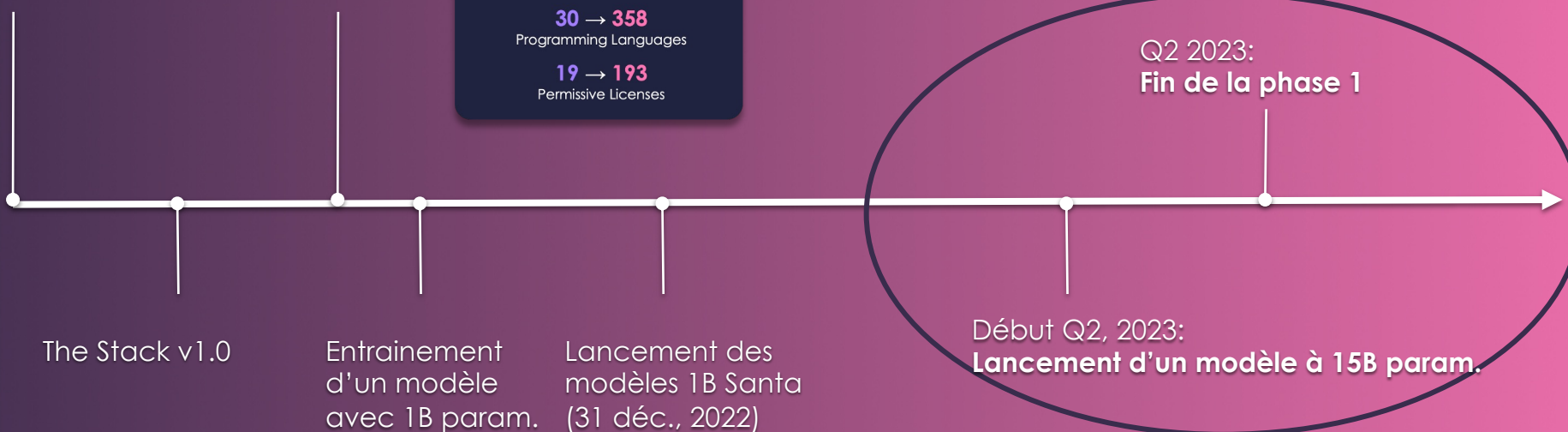
The Stack v1.1
1^{er} déc., 2022

The Stack v1.1

3 TB → 6.4 TB
Volume of data

30 → 358
Programming Languages

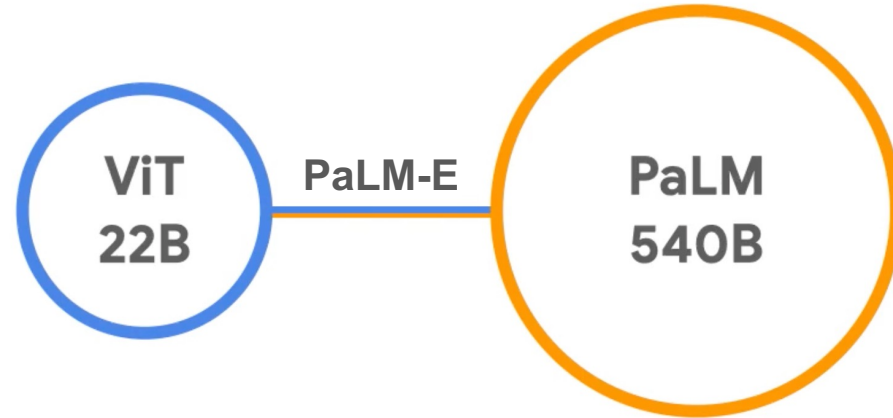
19 → 193
Permissive Licenses



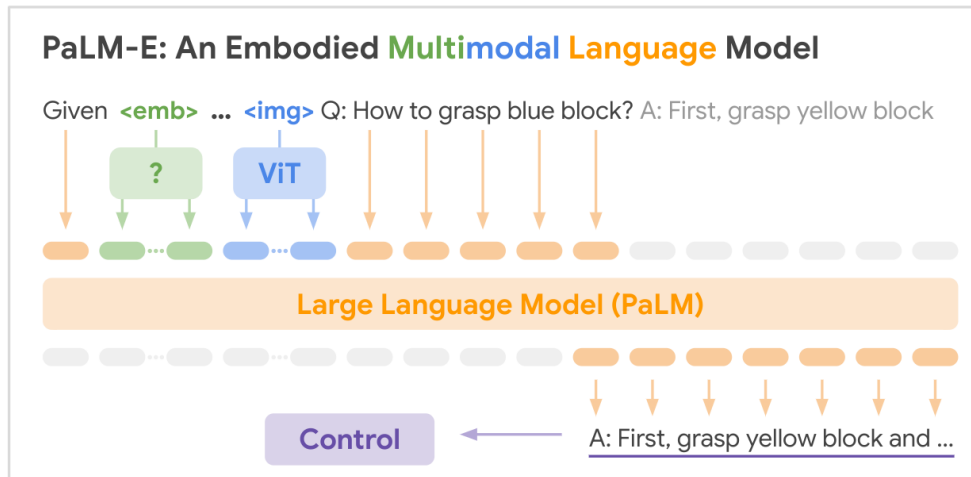
 **bigcode/santacoder**

Plus d'info: <https://www.bigcode-project.org>

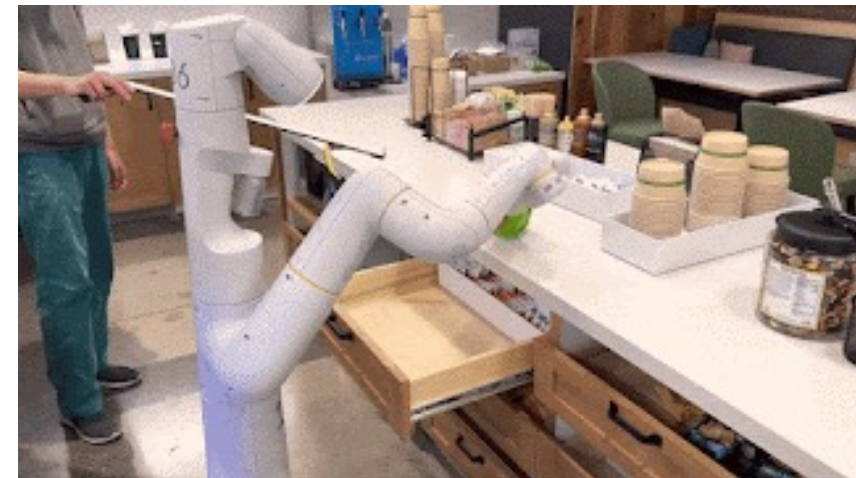
... percevoir, raisonner, planifier et agir



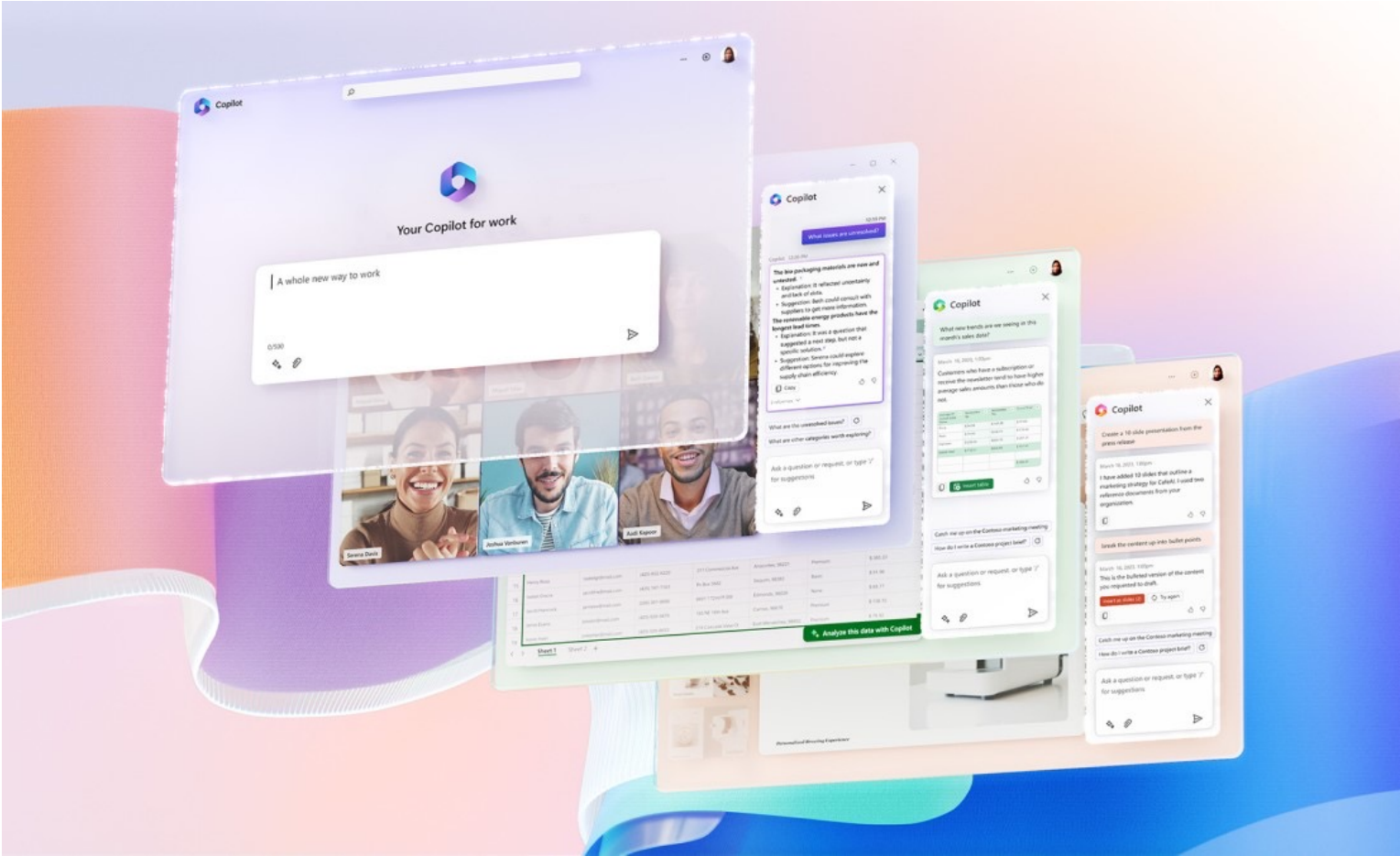
Raisonnement



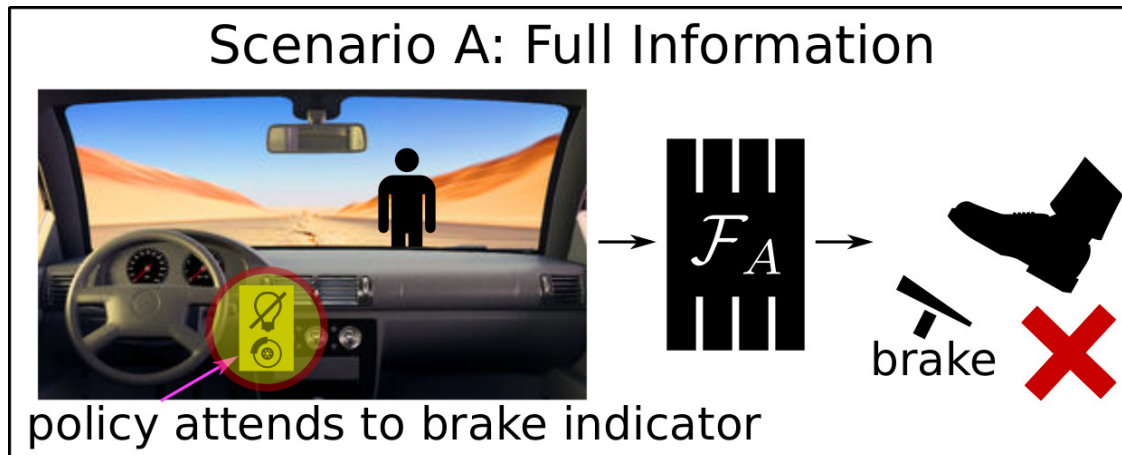
Contrôle



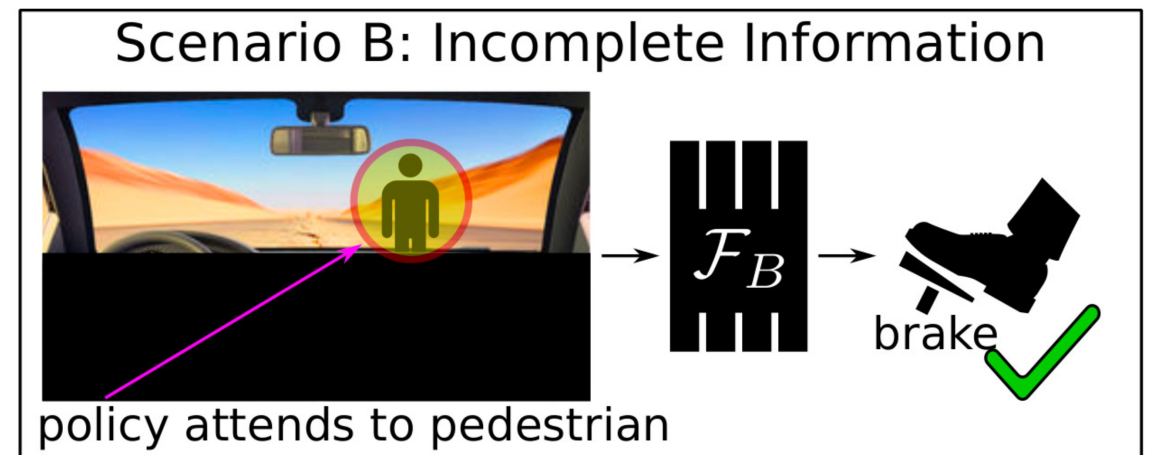
Les copilotes vont changer notre manière de travailler



Malgré tout, ces modèles ne sont que de puissants détecteurs de **corrélations**



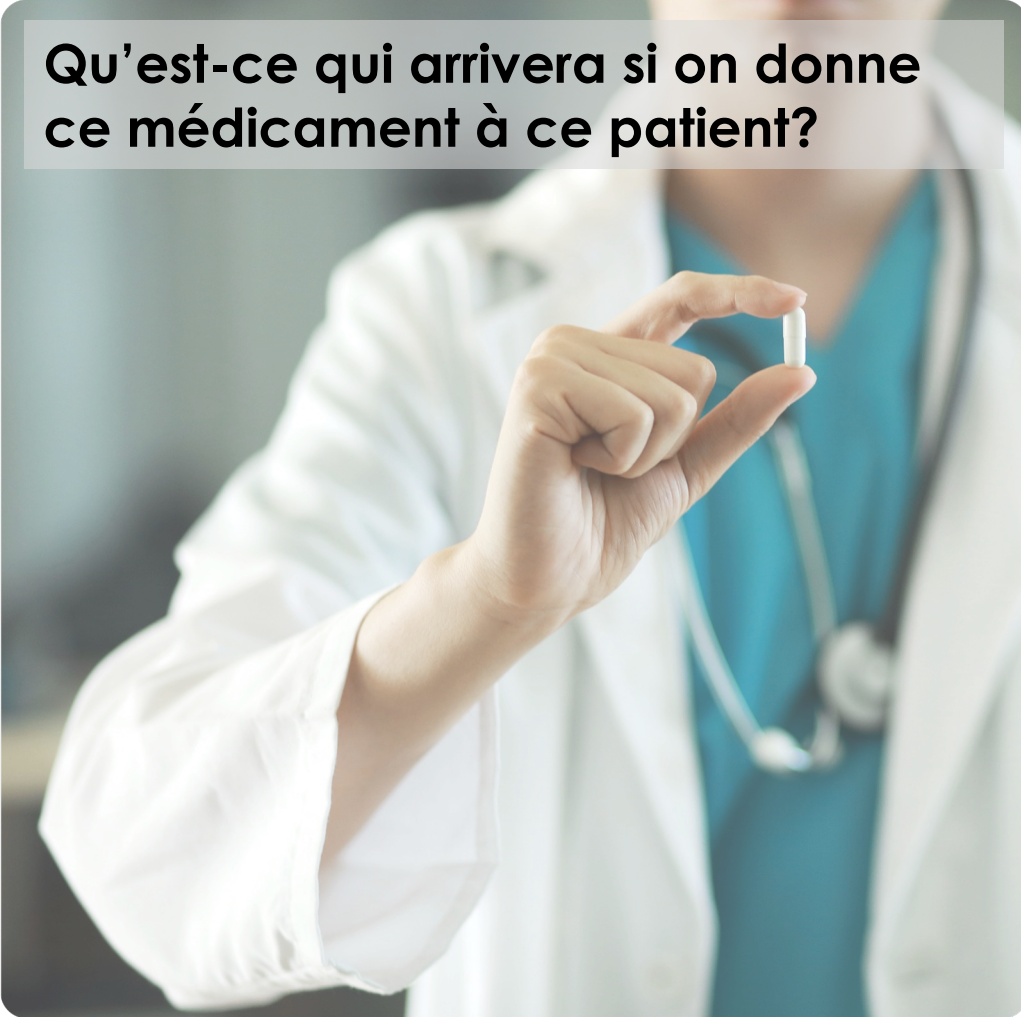
Le modèle croit que l'indicateur lumineux est la cause du freinage.



Le modèle comprend que le piéton est la cause.

Prise de décision: les corrélations ne suffisent pas

Qu'est-ce qui arrivera si on donne ce médicament à ce patient?



Comment vont réagir nos clients si on augmente les prix?



Un autre exemple

Le problème des données historiques



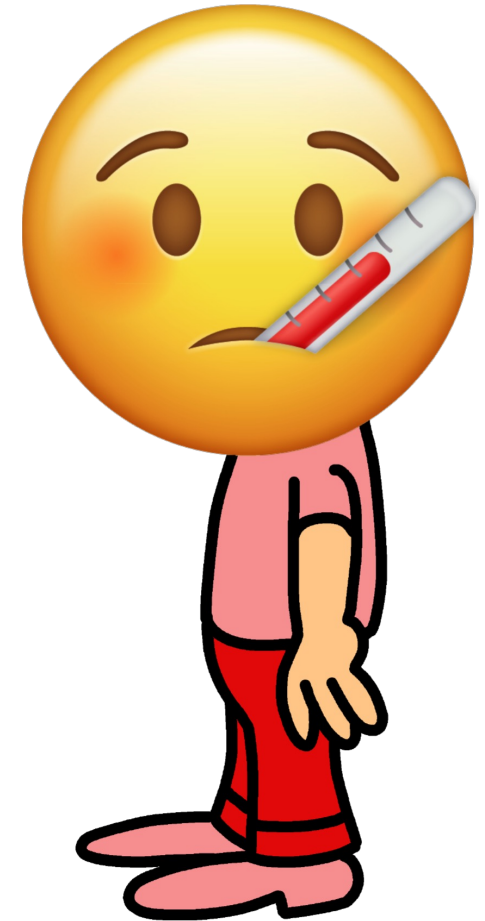
servicenow.

Taux de survie des patients selon le traitement

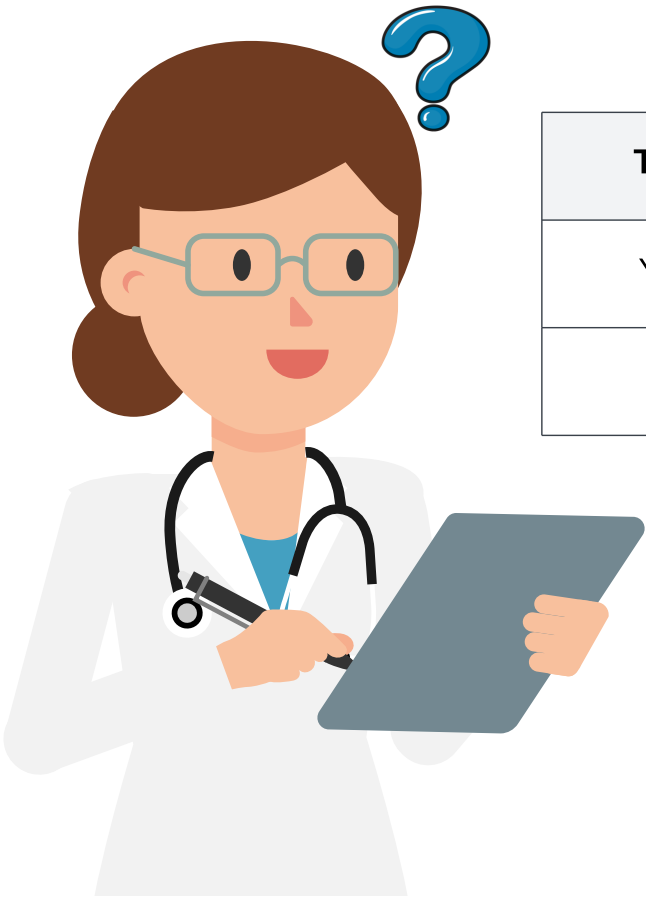
| Treatment | $E[Y A]$ |
|-----------------|----------------------|
| Yes ($A = 1$) | 78% (273/350) |
| No ($A = 0$) | 83% (289/350) |

Conclusion: Le traitement semble **néfaste**

Source: adaptation de [Julious et Mullee \(1994\)](#)



Un autre exemple

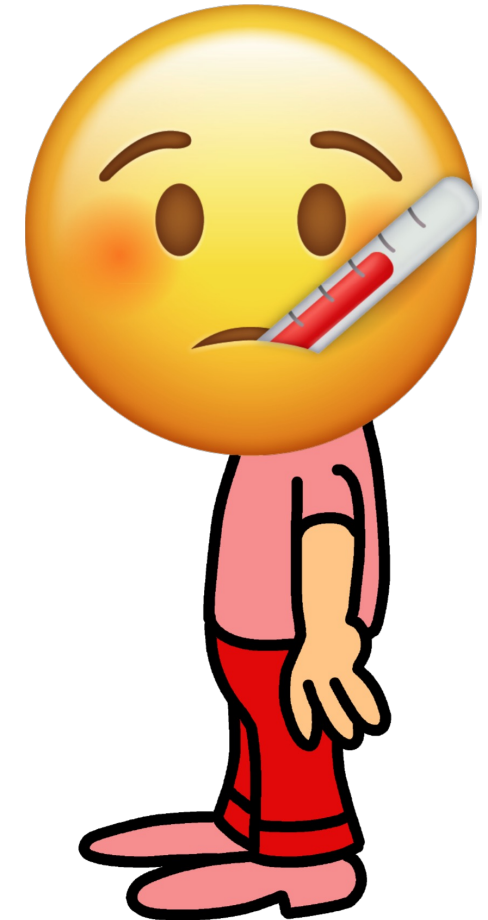


Taux de survie des patients selon le traitement

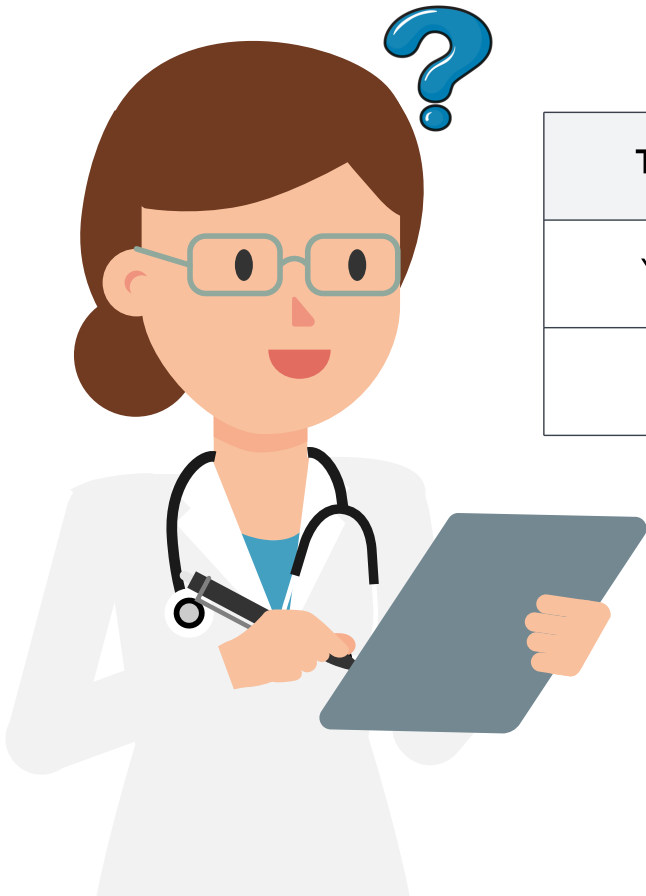
| Treatment | $E[Y A]$ | $E[Y A, Z = mild]$ | $E[Y A, Z = severe]$ |
|-----------------|---------------|----------------------|------------------------|
| Yes ($A = 1$) | 78% (273/350) | 93% (81/87) | 73% (192/263) |
| No ($A = 0$) | 83% (289/350) | 87% (234/270) | 69% (55/80) |

Condition légère: le traitement est **bénéfique**

$$\mathbb{E}[Y | A = 1, Z = mild] - \mathbb{E}[Y | A = 0, Z = mild] = 6\%$$



Un autre exemple



Taux de survie des patients selon le traitement

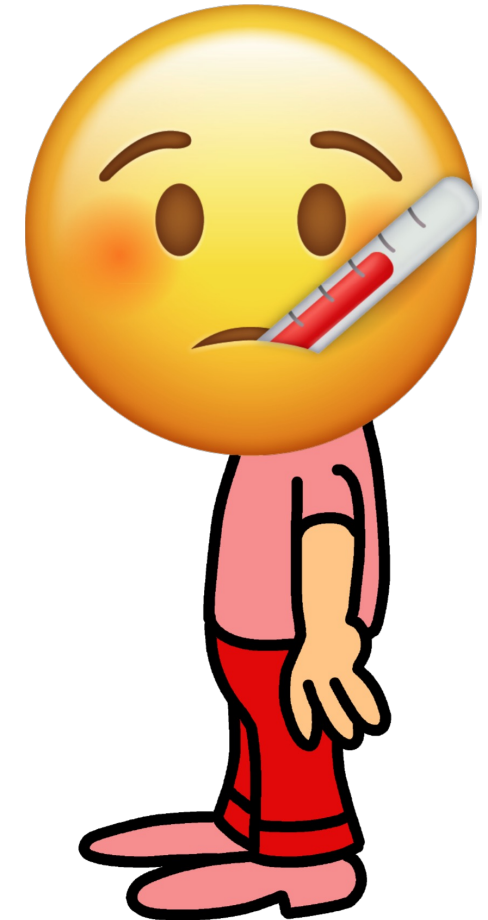
| Treatment | $\mathbb{E}[Y A]$ | $\mathbb{E}[Y A, Z = \text{mild}]$ | $\mathbb{E}[Y A, Z = \text{severe}]$ |
|-----------------|---------------------|--------------------------------------|----------------------------------------|
| Yes ($A = 1$) | 78% (273/350) | 93% (81/87) | 73% (192/263) |
| No ($A = 0$) | 83% (289/350) | 87% (234/270) | 69% (55/80) |

Condition légère: le traitement est **bénéfique**

$$\mathbb{E}[Y | A = 1, Z = \text{mild}] - \mathbb{E}[Y | A = 0, Z = \text{mild}] = 6\%$$

Condition sévère: le traitement est **bénéfique**

$$\mathbb{E}[Y | A = 1, Z = \text{severe}] - \mathbb{E}[Y | A = 0, Z = \text{severe}] = 4\%$$



Pourquoi?

Les données historiques reflètent un protocole



Pourquoi?

Les données historiques reflètent un protocole



Z = mild



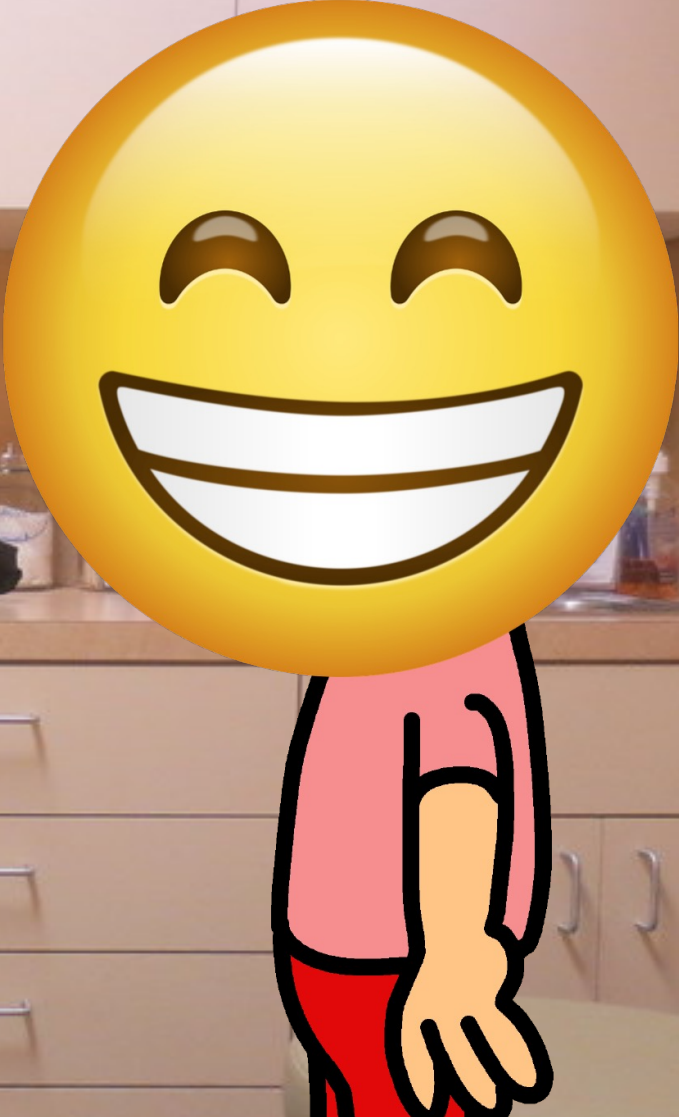
Pourquoi?

Les données historiques reflètent un protocole



Vous allez survivre. $A = 0$

Z = mild



Pourquoi?

Les données historiques reflètent un protocole

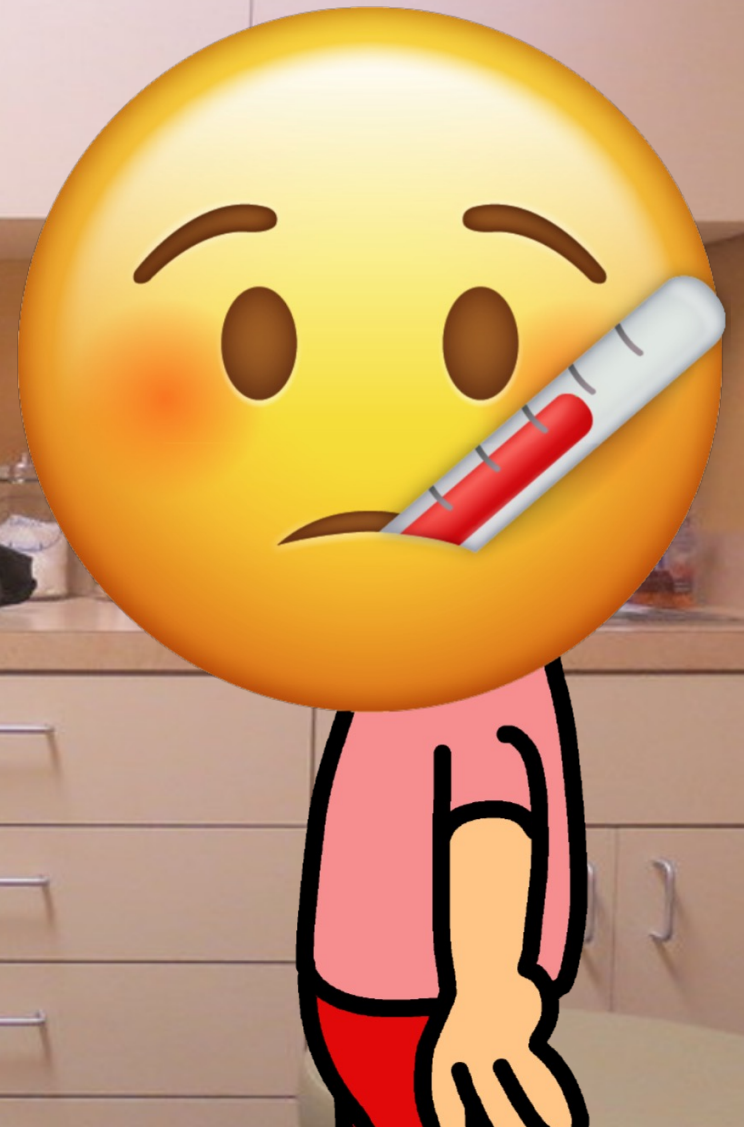


Pourquoi?

Les données historiques reflètent un protocole



Z = severe



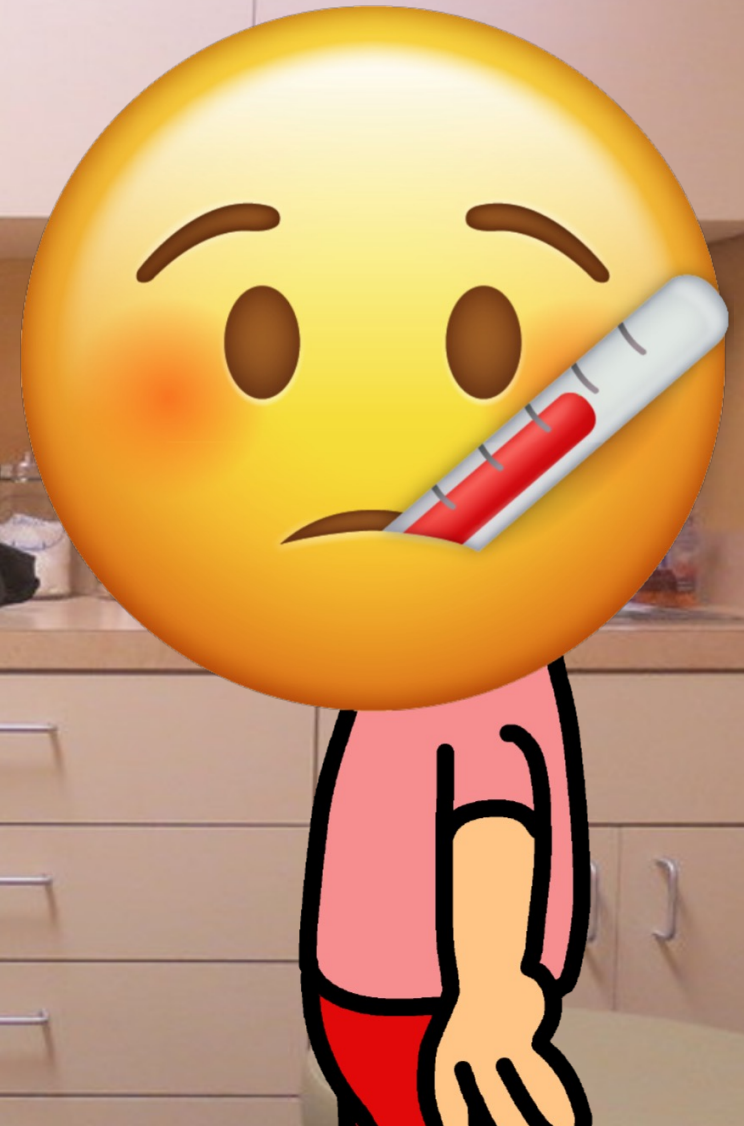
Pourquoi?

Les données historiques reflètent un protocole



Vous êtes en danger! A = 1

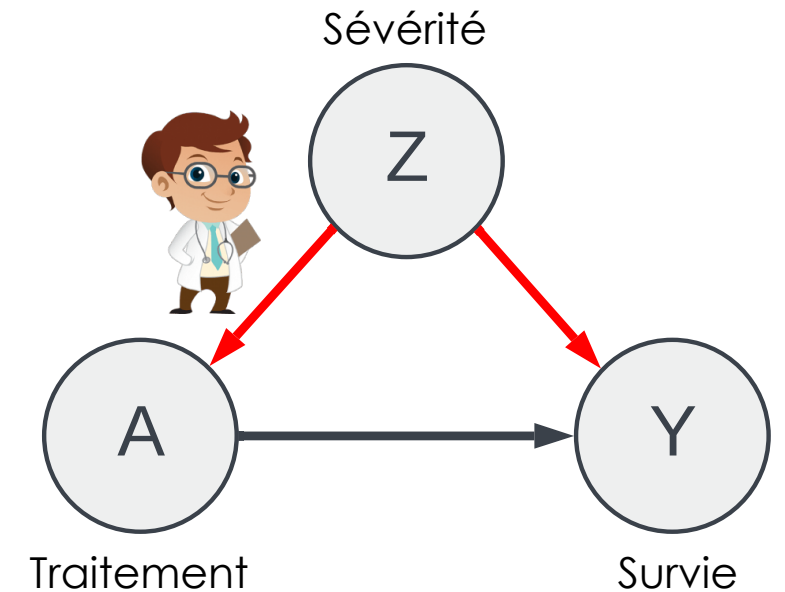
Z = severe



Le paradoxe de Simpson

| Treatment | $E[Y A]$ | $E[Y A, Z = mild]$ | $E[Y A, Z = severe]$ |
|-----------------|----------------------|----------------------|------------------------|
| Yes ($A = 1$) | 78% (273/350) | 93% (81/87) | 73% (192/263) |
| No ($A = 0$) | 83% (289/350) | 87% (234/270) | 69% (55/80) |

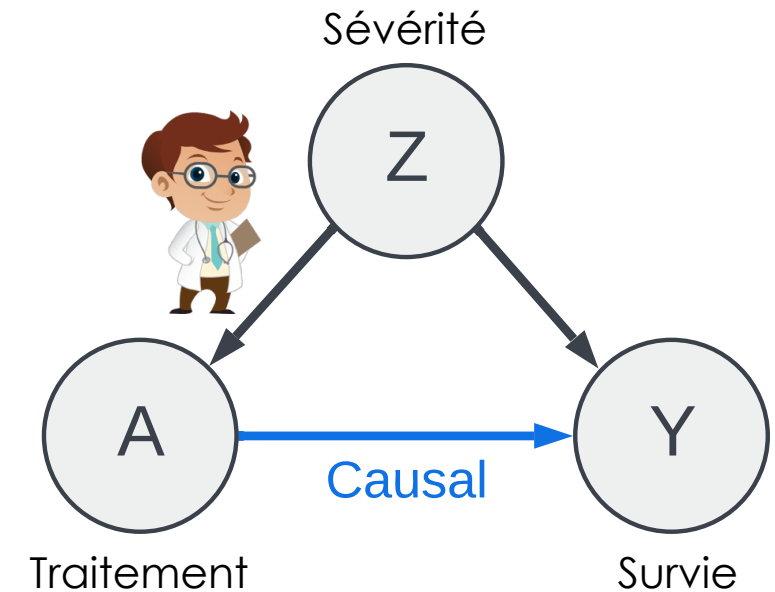
Confusion: Les patients **les plus malades** étaient plus susceptibles de **recevoir le traitement** et aussi de **ne pas survivre!**



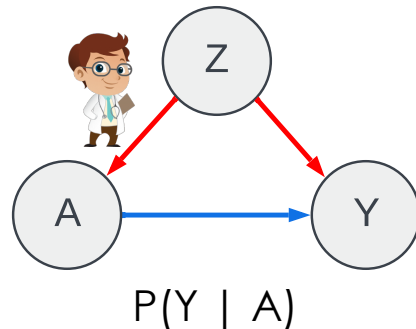
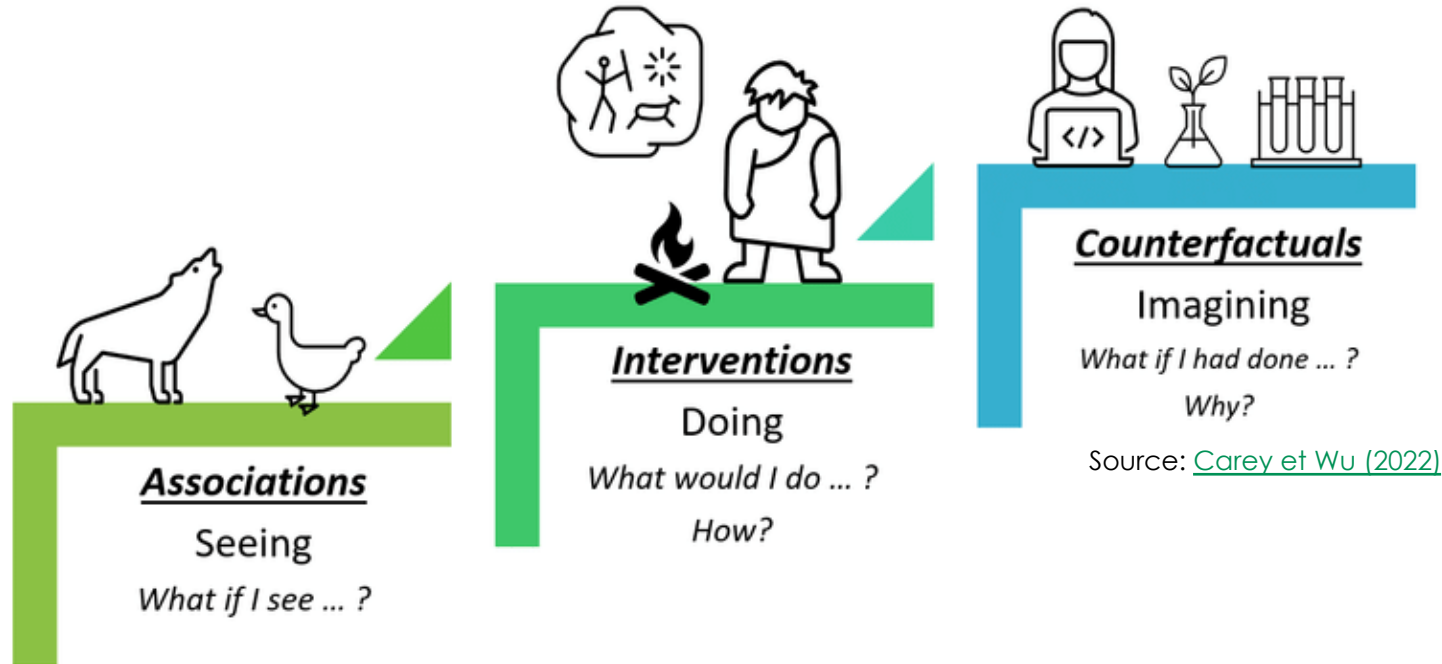
Le paradoxe de Simpson

| Treatment | $E[Y A]$ | $E[Y A, Z = mild]$ | $E[Y A, Z = severe]$ |
|-----------------|----------------------|----------------------|------------------------|
| Yes ($A = 1$) | 78% (273/350) | 93% (81/87) | 73% (192/263) |
| No ($A = 0$) | 83% (289/350) | 87% (234/270) | 69% (55/80) |

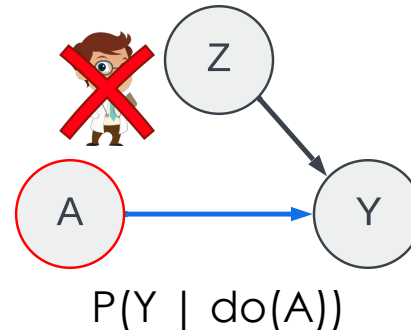
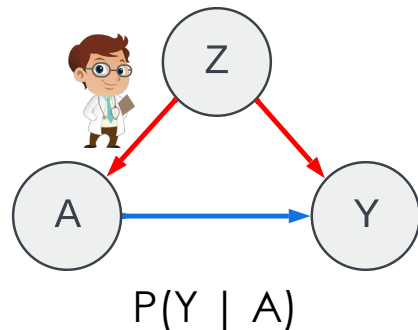
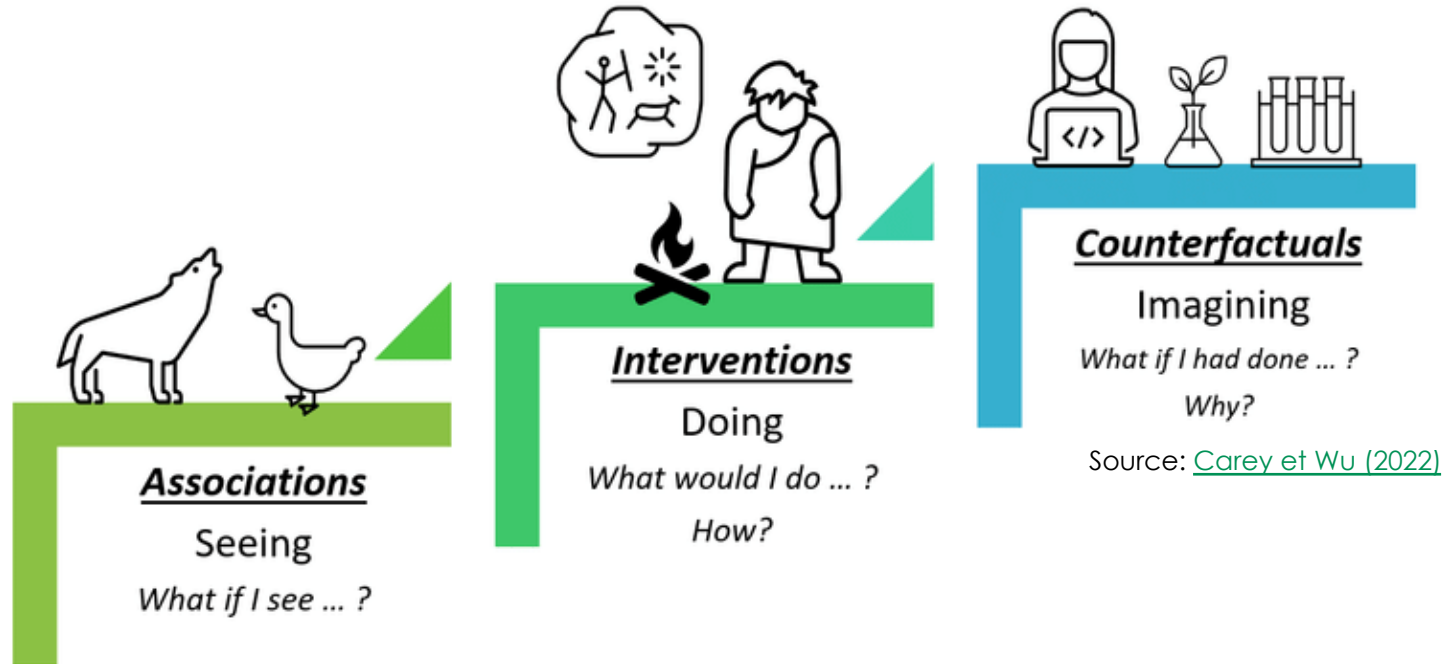
Confusion: Les patients **les plus malades** étaient plus susceptibles de **recevoir le traitement** et aussi de **ne pas survivre!**



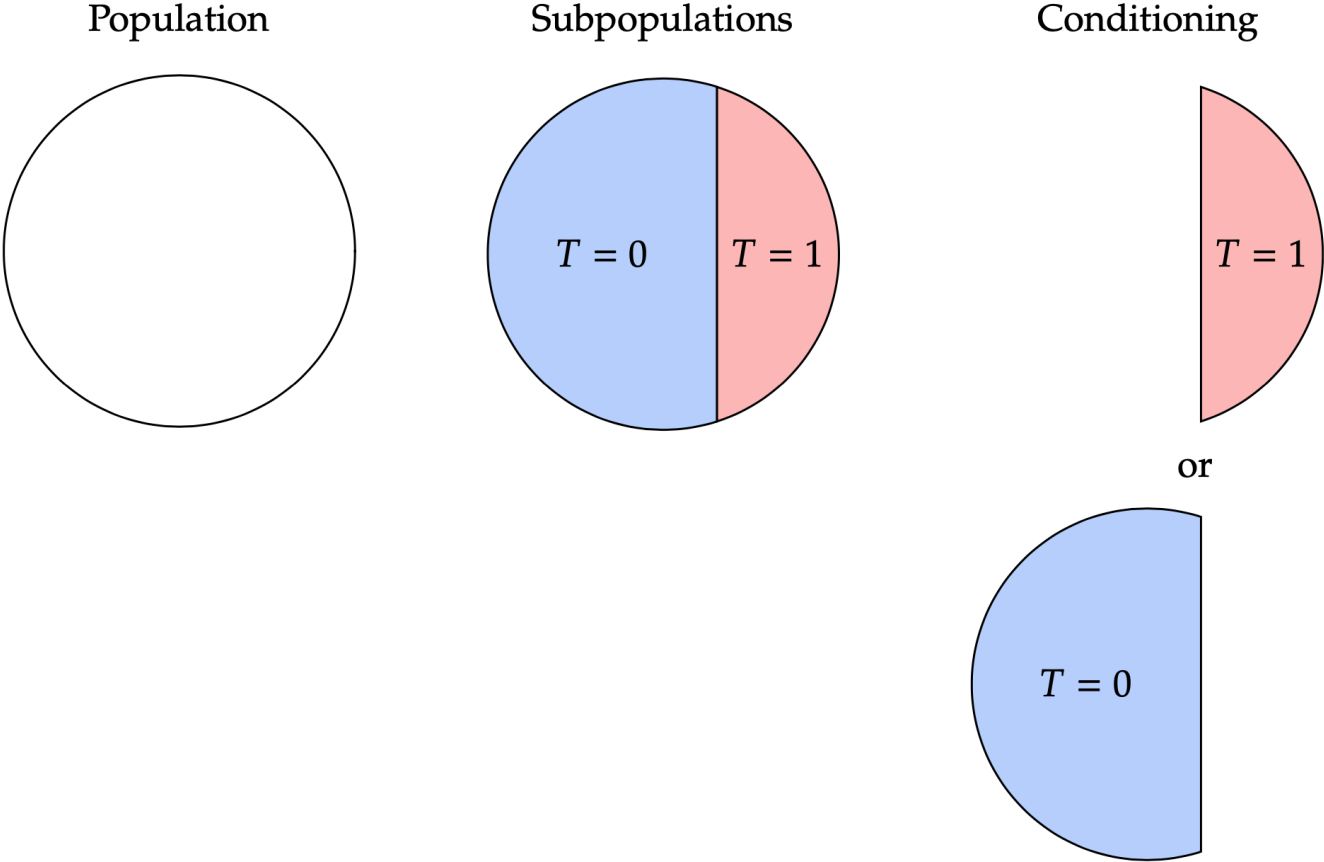
Toutes les questions ne sont pas égales



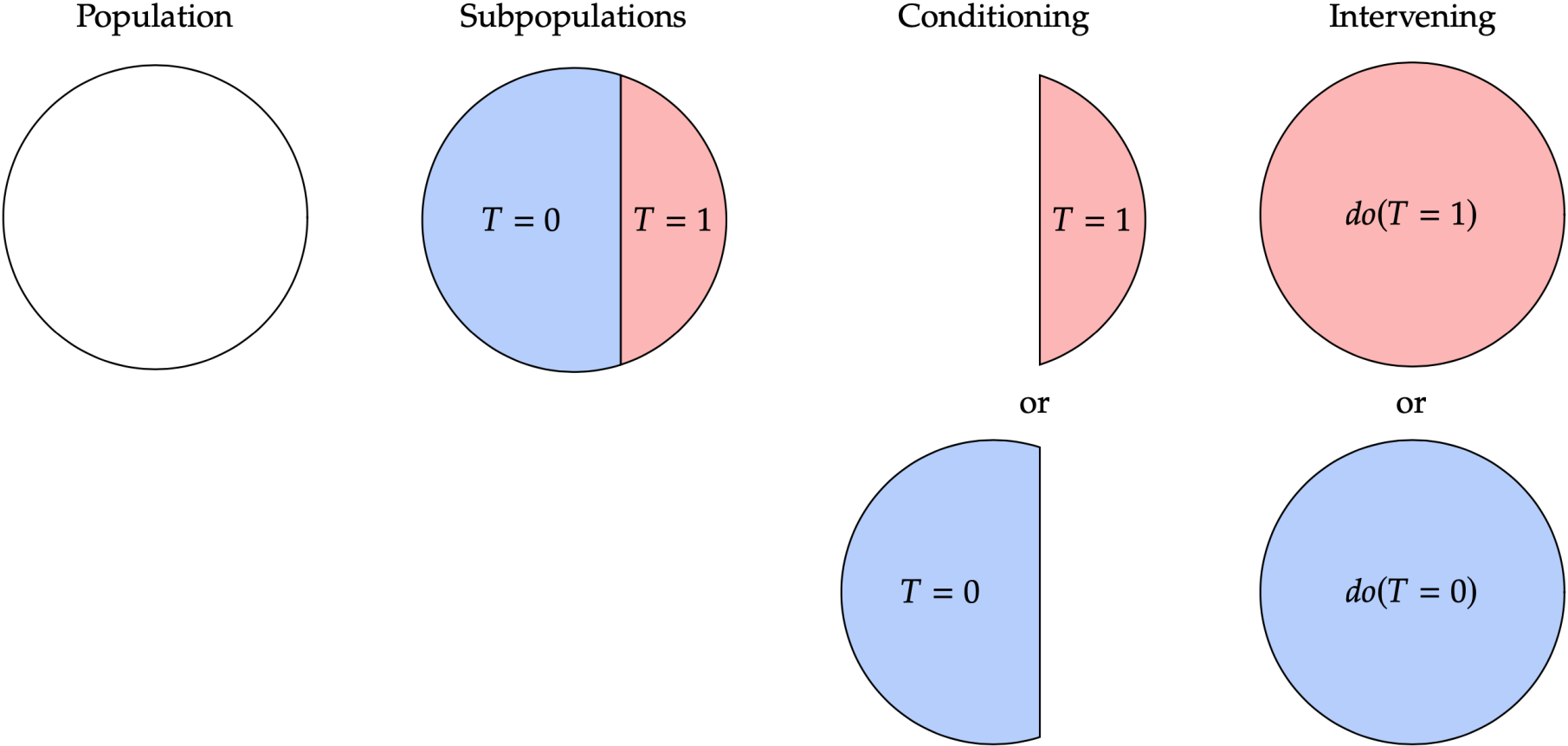
Toutes les questions ne sont pas égales



Association vs intervention

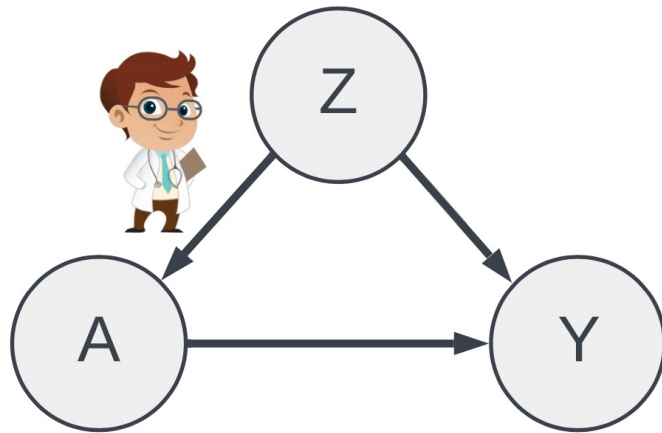


Association vs intervention



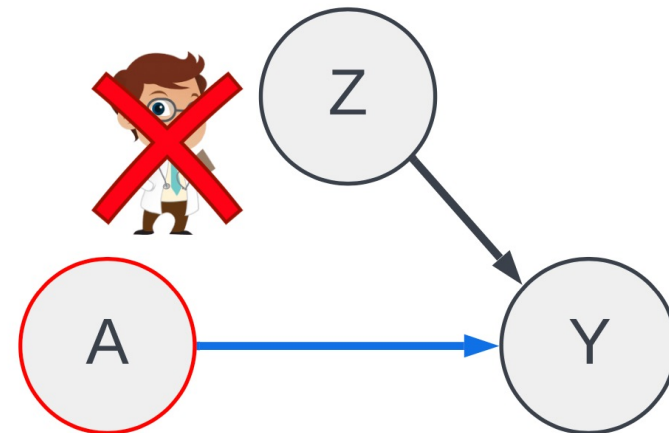
Cadre théorique: Réseaux Bayésiens Causaux

Observations



$$P(a, y, z) = P(z) P(a | z) P(y | a, z)$$

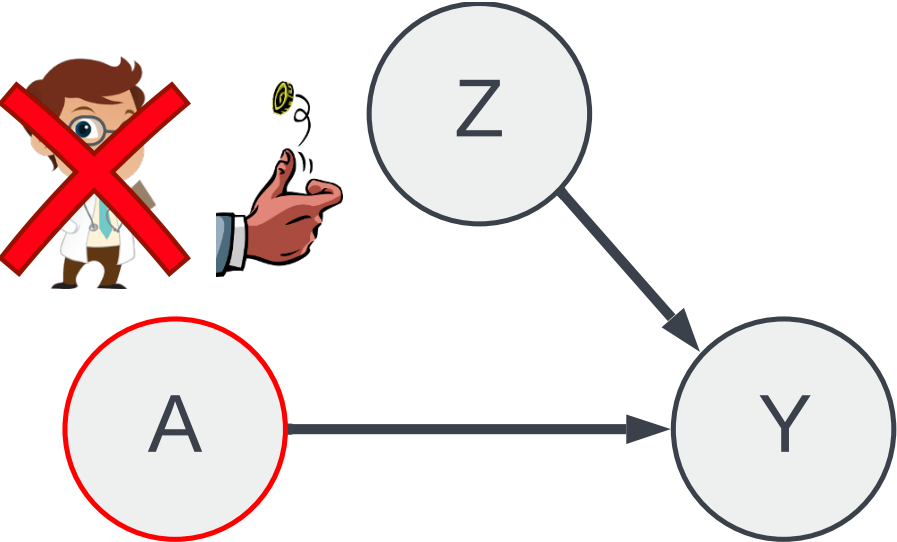
Intervention



$$P'(a, y, z) = P(z) P'(a) P(y | a, z)$$

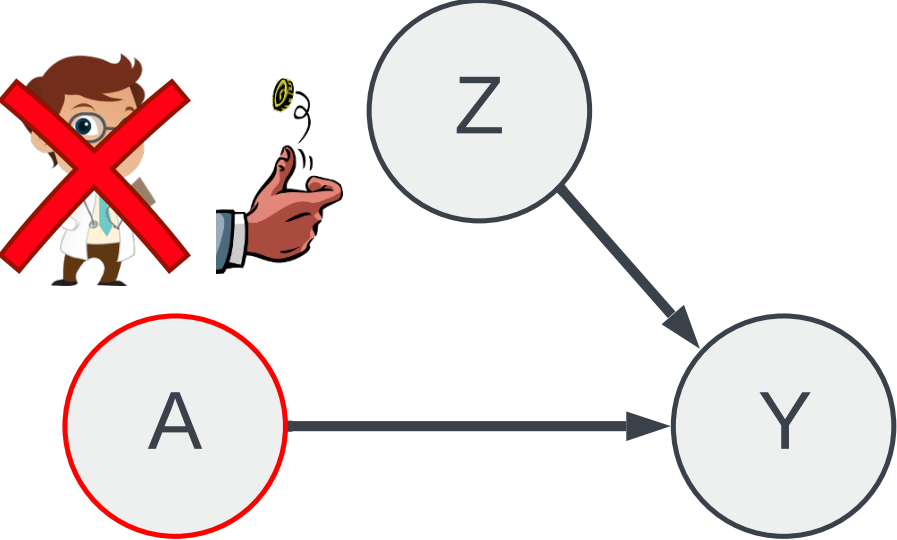
La **distribution** de chaque **variable** est exprimée en fonction de ses **parents** dans le **graphe causal**

Expériences randomisées



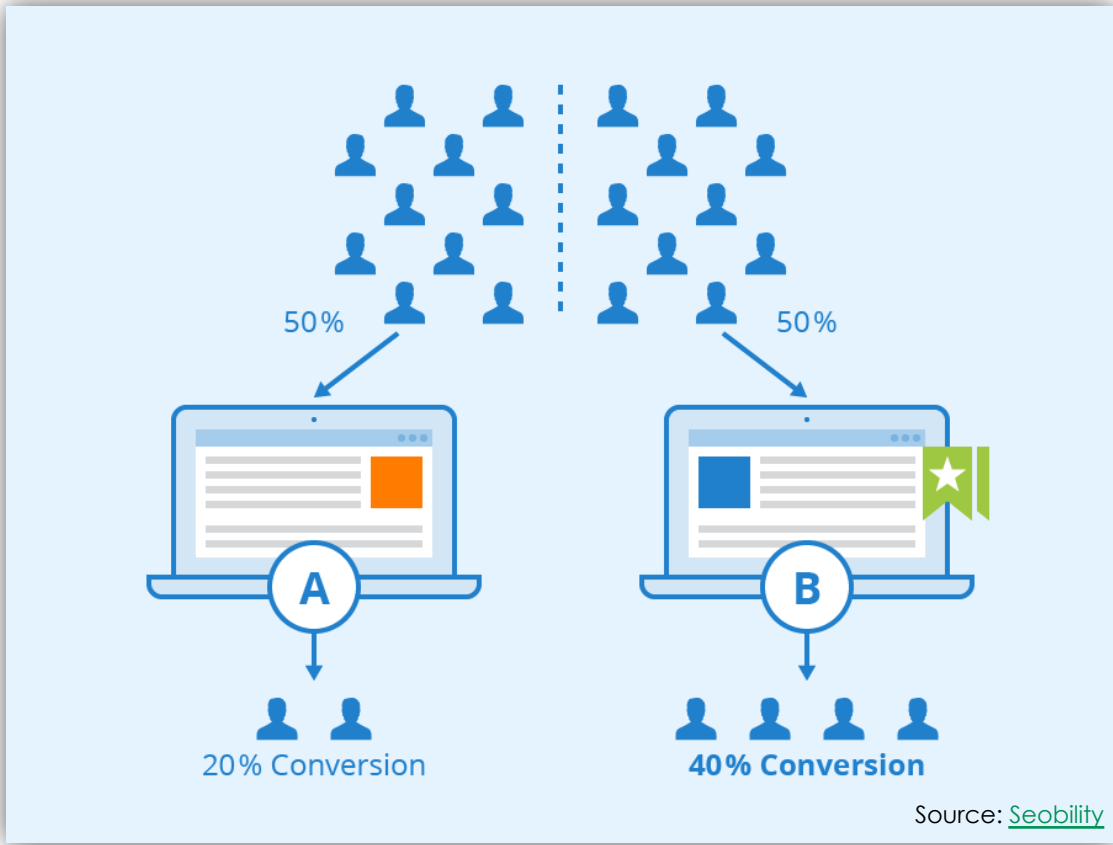
$$P'(a, y, z) = P(z) P'(a) P(y | a, z)$$

Expériences randomisées



$$P'(a, y, z) = P(z) P'(a) P(y | a, z)$$

A/B testing

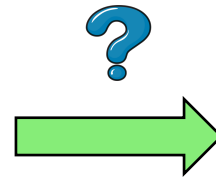
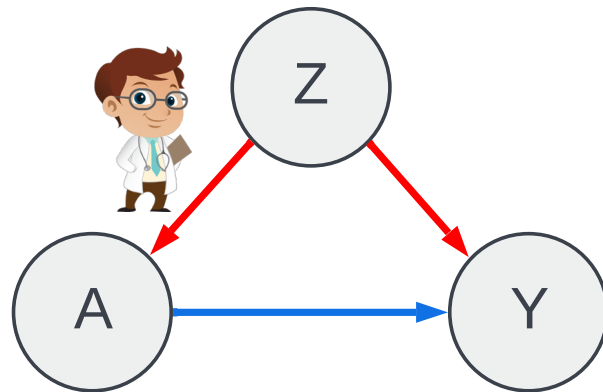


Source: [Seobility](#)

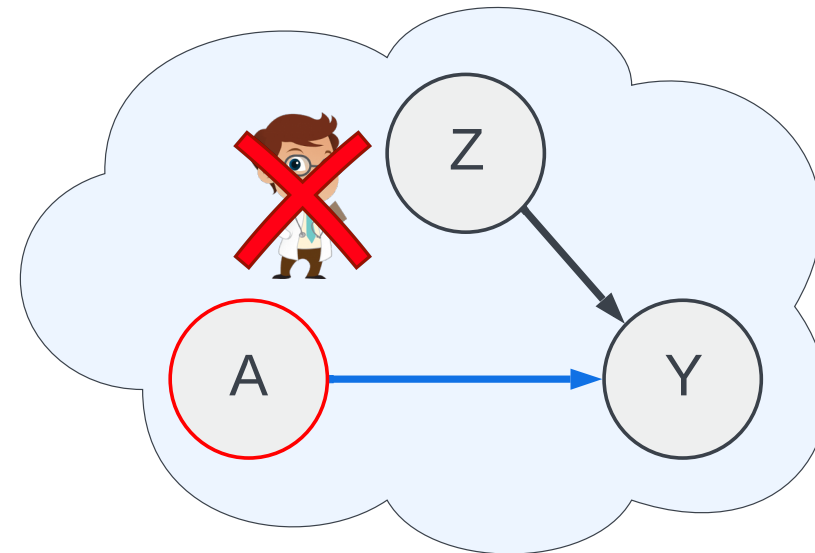
Tout ne peut pas être randomisé

Pouvons-nous tout de même estimer l'effet d'interventions?

Distribution observée



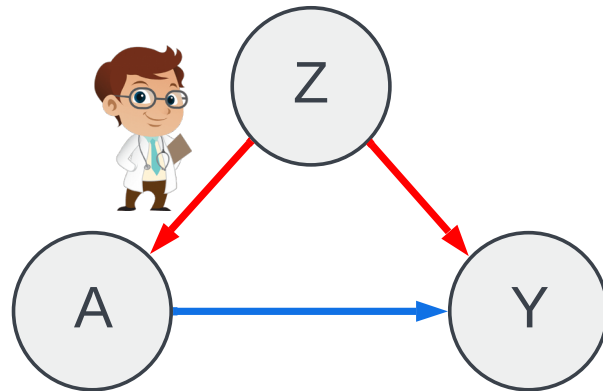
Distribution d'intervention



Tout ne peut pas être randomisé

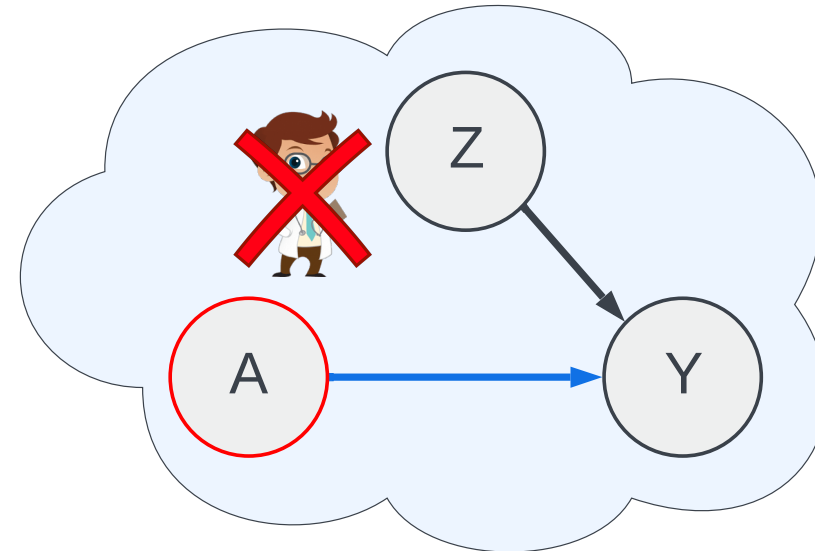
Pouvons-nous tout de même estimer l'effet d'interventions?

Distribution observée



Do-calculus
[Pearl \(1995\)](#)

Distribution d'intervention

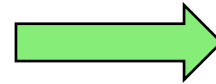
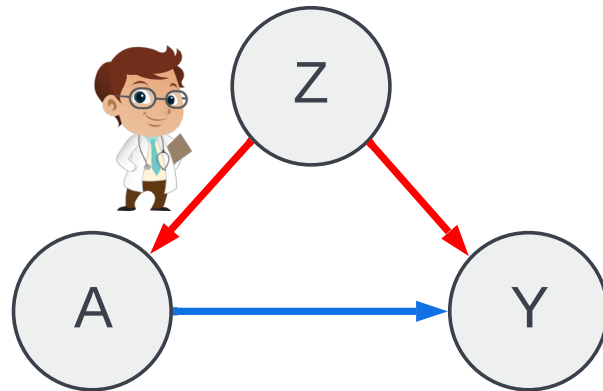


Identification causale: Il est parfois possible de traduire une **quantité causale** $P(Y \mid \text{do}(A))$ en **quantité observationnelle** (sans opérateur do).

Tout ne peut pas être randomisé

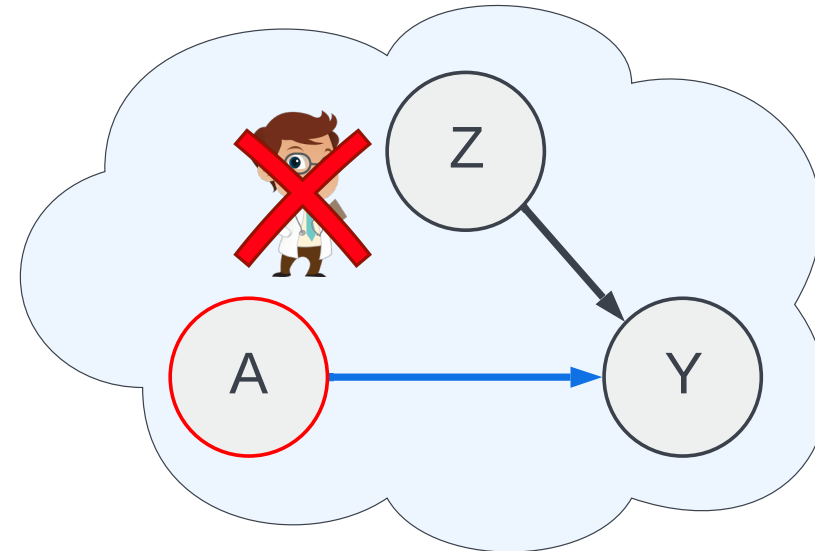
Pouvons-nous tout de même estimer l'effet d'interventions?

Distribution observée



Do-calculus
[Pearl \(1995\)](#)

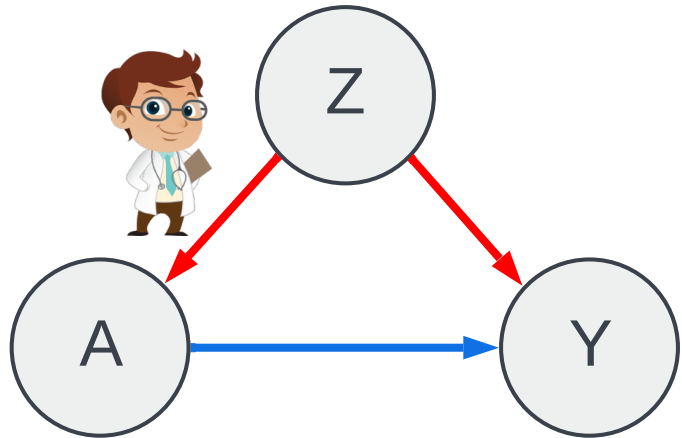
Distribution d'intervention



Identification causale: Il est parfois possible de traduire une **quantité causale** $P(Y \mid \text{do}(A))$ en **quantité observationnelle** (sans opérateur do).

→ Graphe + quantité interv. → Solution / Impossible (temps polynomial)

Un exemple d'identification causale



$$P(y \mid do(A = a')) = \sum_a \sum_z P(a, y, z \mid do(A = a'))$$

< Truncated factorization >

$$= \sum_a \sum_z P(y \mid a', z) \cdot P(z) \cdot \delta_{a=a'}$$

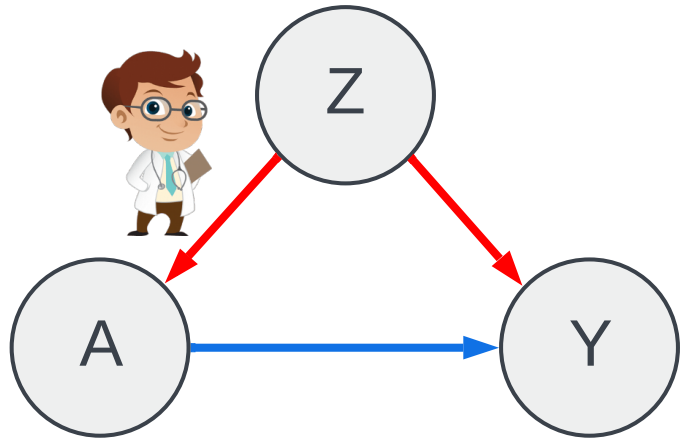
< Density is zero for all $a \neq a'$ >

$$= 0 + \sum_z P(y \mid a', z) \cdot P(z) \cdot 1$$

< Cleaning up a bit >

$$= \sum_z P(y \mid a', z) \cdot P(z)$$

Un exemple d'identification causale



$$P(y \mid do(A = a')) = \sum_a \sum_z P(a, y, z \mid do(A = a'))$$

< Truncated factorization >

$$= \sum_a \sum_z P(y \mid a', z) \cdot P(z) \cdot \delta_{a=a'}$$

< Density is zero for all $a \neq a'$ >

$$= 0 + \sum_z P(y \mid a', z) \cdot P(z) \cdot 1$$

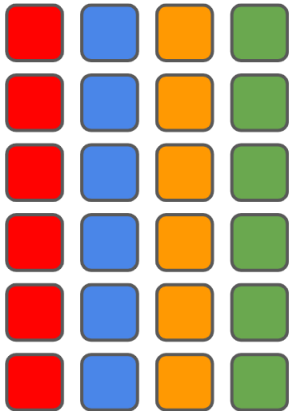
< Cleaning up a bit >

$$= \sum_z P(y \mid a', z) \cdot P(z)$$

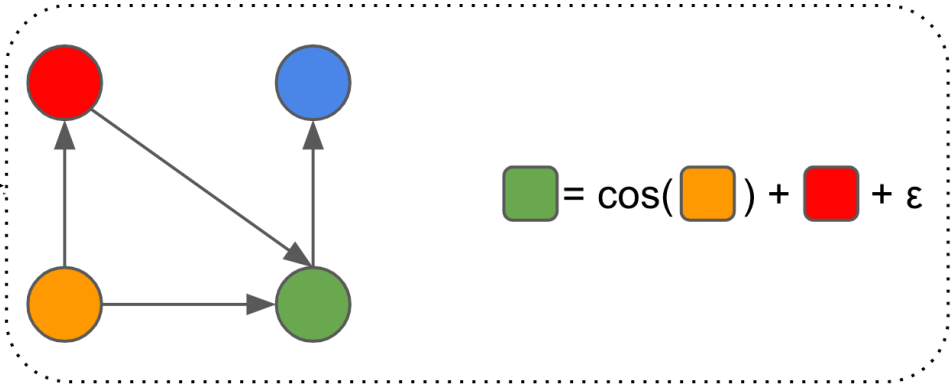
Limitation: Il faut connaître le graphe

Découverte de structures causales

Problème:



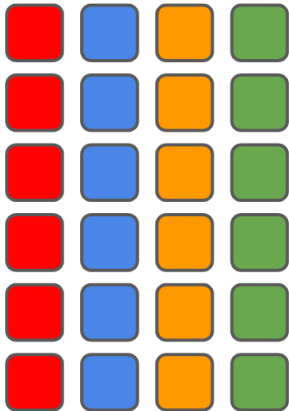
Ensemble d'observations



Graphe causal et équations structurelles

Découverte de structures causales

Problème:



Ensemble d'observation

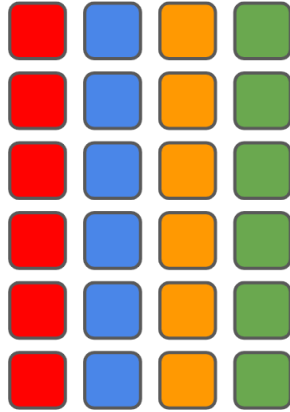
| p | number of DAGs with p nodes |
|-----|---------------------------------------------------------------------------|
| 1 | 1 |
| 2 | 3 |
| 3 | 25 |
| 4 | 543 |
| 5 | 29281 |
| 6 | 3781503 |
| 7 | 1138779265 |
| 8 | 783702329343 |
| 9 | 1213442454842881 |
| 10 | 4175098976430598143 |
| 11 | 31603459396418917607425 |
| 12 | 521939651343829405020504063 |
| 13 | 18676600744432035186664816926721 |
| 14 | 1439428141044398334941790719839535103 |
| 15 | 237725265553410354992180218286376719253505 |
| 16 | 83756670773733320287699303047996412235223138303 |
| 17 | 62707921196923889899446452602494921906963551482675201 |
| 18 | 99421195322159515895228914592354524516555026878588305014783 |
| 19 | 332771901227107591736177573311261125883583076258421902583546773505 |
| 20 | 2344880451051088988152559855229099188899081192234291298795803236068491263 |

$$\text{Green square} = \cos(\text{Orange square}) + \text{Red square} + \epsilon$$

relations structurelles

Découverte de structures causales

Problème:



Ensemble d'observation

| p | number of DAGs with p nodes |
|-----|---------------------------------------------------------------------------|
| 1 | 1 |
| 2 | 3 |
| 3 | 25 |
| 4 | 543 |
| 5 | 29281 |
| 6 | 3781503 |
| 7 | 1138779265 |
| 8 | 783702329343 |
| 9 | 1213442454842881 |
| 10 | 4175098976430598143 |
| 11 | 31603459396418917607425 |
| 12 | 521939651343829405020504063 |
| 13 | 18676600744432035186664816926721 |
| 14 | 1439428141044398334941790719839535103 |
| 15 | 237725265553410354992180218286376719253505 |
| 16 | 83756670773733320287699303047996412235223138303 |
| 17 | 62707921196923889899446452602494921906963551482675201 |
| 18 | 99421195322159515895228914592354524516555026878588305014783 |
| 19 | 332771901227107591736177573311261125883583076258421902583546773505 |
| 20 | 2344880451051088988152559855229099188899081192234291298795803236068491263 |

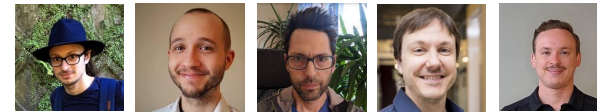
$$\text{Green square} = \cos(\text{Orange square}) + \text{Red square} + \epsilon$$

relations structurelles

État de l'art:

- Indépendances conditionnelles (Spirtes et al., 2000)
- Heuristiques et énumération (Chickering 2002)
- Optimisation continue (Zheng et al., 2018) ★

Differentiable Causal Discovery from Interventional Data



Philippe Brouillard*
Mila, Université de Montréal

Sébastien Lachapelle*
Mila, Université de Montréal

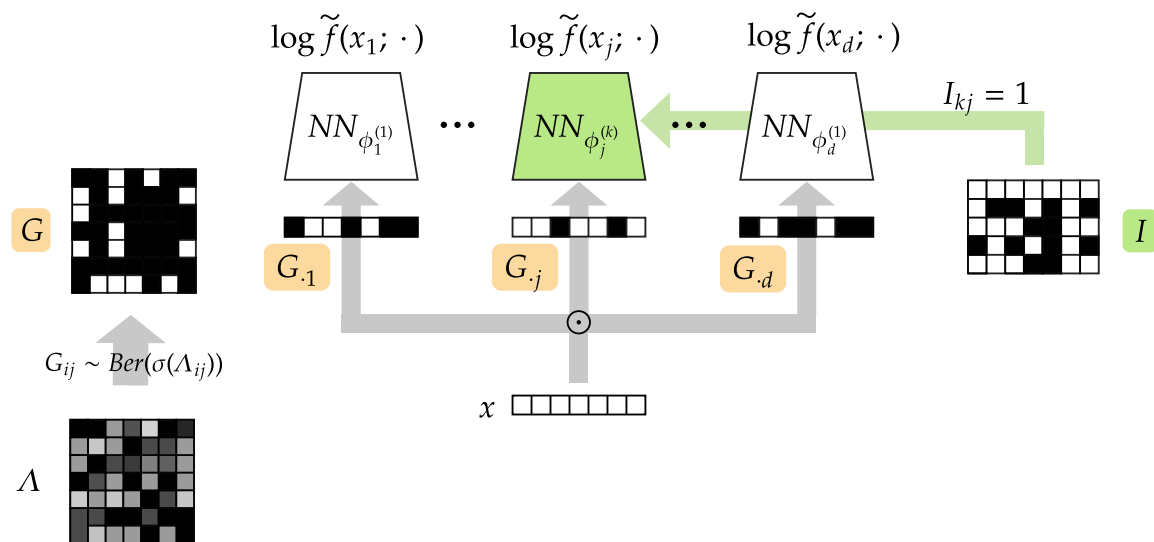
Alexandre Lacoste
Element AI

Simon Lacoste-Julien
Mila, Université de Montréal
Canada CIFAR AI Chair

Alexandre Drouin
Element AI

<https://arxiv.org/abs/2007.01754>

Modèle:



Entraînement:

$$\sup_{\phi} \mathbb{E}_{G \sim \sigma(\Lambda)} \left[\sum_{k=1}^K \mathbb{E}_{X \sim p^{(k)}} \log f^{(k)}(X; G, I^*, \phi) - \lambda \|G\|_0 \right] \quad s.t. \quad \underbrace{\text{Tr}(e^{\sigma(\Lambda)}) - d = 0}_{\text{Acyclicity constraint}}$$

Bien modéliser les données

Differentiable Causal Discovery from Interventional Data



Philippe Brouillard*
Mila, Université de Montréal

Sébastien Lachapelle*
Mila, Université de Montréal

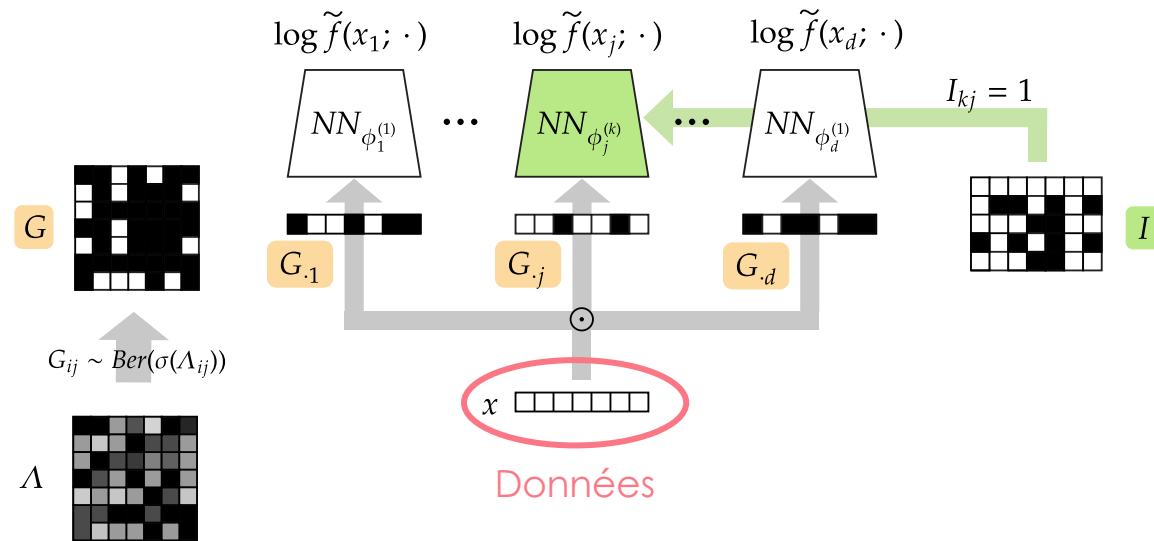
Alexandre Lacoste
Element AI

Simon Lacoste-Julien
Mila, Université de Montréal
Canada CIFAR AI Chair

Alexandre Drouin
Element AI

<https://arxiv.org/abs/2007.01754>

Modèle:



Entraînement:

$$\sup_{\phi} \mathbb{E}_{G \sim \sigma(\Lambda)} \left[\sum_{k=1}^K \mathbb{E}_{X \sim p^{(k)}} \log f^{(k)}(X; G, I^*, \phi) - \lambda \|G\|_0 \right] \quad s.t. \quad \underbrace{\text{Tr} \left(e^{\sigma(\Lambda)} \right) - d = 0}_{\text{Acyclicity constraint}}$$

Bien modéliser les données

Differentiable Causal Discovery from Interventional Data



Philippe Brouillard*
Mila, Université de Montréal

Sébastien Lachapelle*
Mila, Université de Montréal

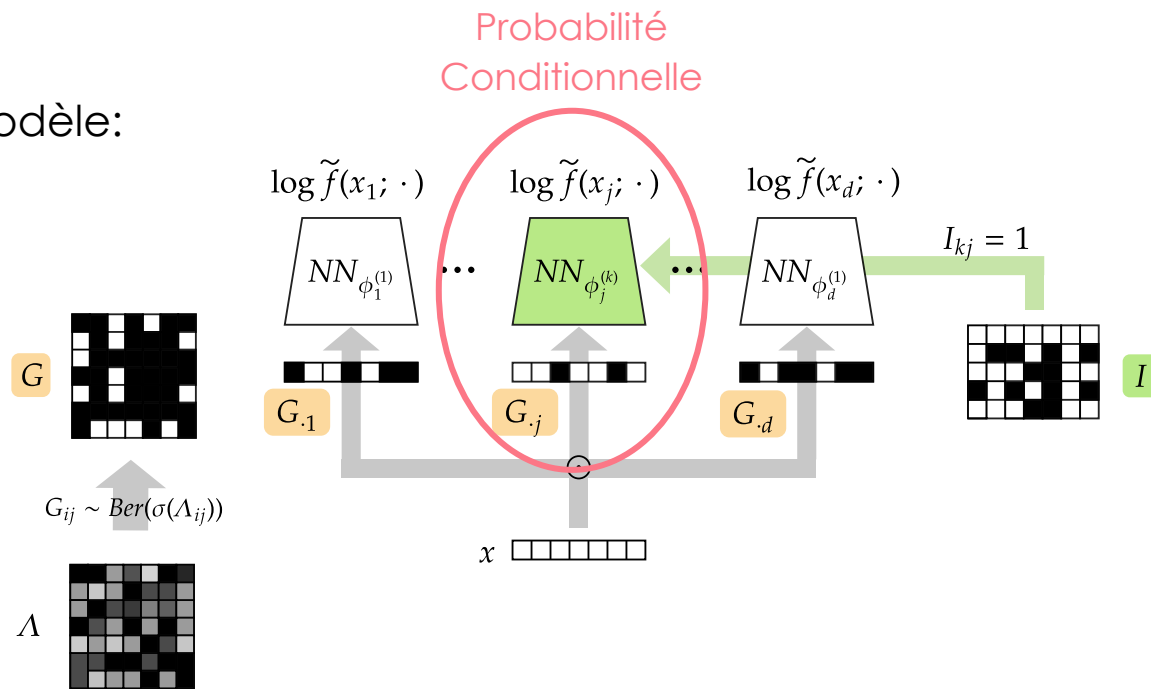
Alexandre Lacoste
Element AI

Simon Lacoste-Julien
Mila, Université de Montréal
Canada CIFAR AI Chair

Alexandre Drouin
Element AI

<https://arxiv.org/abs/2007.01754>

Modèle:



Entraînement:

$$\sup_{\phi} \mathbb{E}_{G \sim \sigma(\Lambda)} \left[\sum_{k=1}^K \mathbb{E}_{X \sim p^{(k)}} \log f^{(k)}(X; G, I^*, \phi) - \lambda \|G\|_0 \right] \quad s.t. \quad \underbrace{\text{Tr} \left(e^{\sigma(\Lambda)} \right) - d = 0}_{\text{Acyclicity constraint}}$$

Bien modéliser les données

Differentiable Causal Discovery from Interventional Data



Philippe Brouillard*
Mila, Université de Montréal

Sébastien Lachapelle*
Mila, Université de Montréal

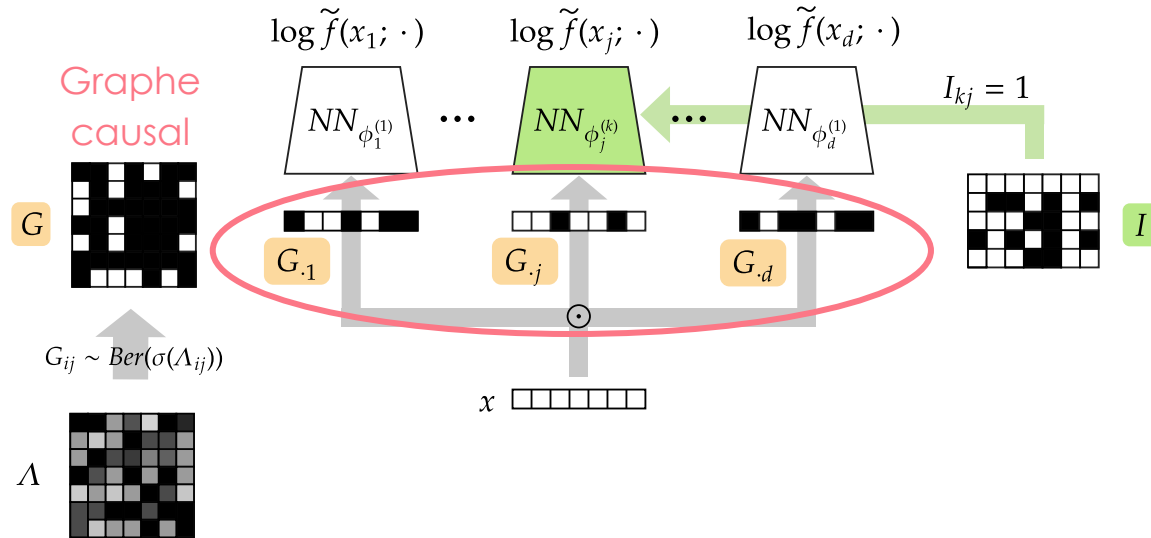
Alexandre Lacoste
Element AI

Simon Lacoste-Julien
Mila, Université de Montréal
Canada CIFAR AI Chair

Alexandre Drouin
Element AI

<https://arxiv.org/abs/2007.01754>

Modèle:



Entraînement:

$$\sup_{\phi} \mathbb{E}_{G \sim \sigma(\Lambda)} \left[\sum_{k=1}^K \mathbb{E}_{X \sim p^{(k)}} \log f^{(k)}(X; G, I^*, \phi) - \lambda \|G\|_0 \right] \quad s.t. \quad \underbrace{\text{Tr} \left(e^{\sigma(\Lambda)} \right) - d = 0}_{\text{Acyclicity constraint}}$$

Bien modéliser les données

Differentiable Causal Discovery from Interventional Data



Philippe Brouillard*
Mila, Université de Montréal

Sébastien Lachapelle*
Mila, Université de Montréal

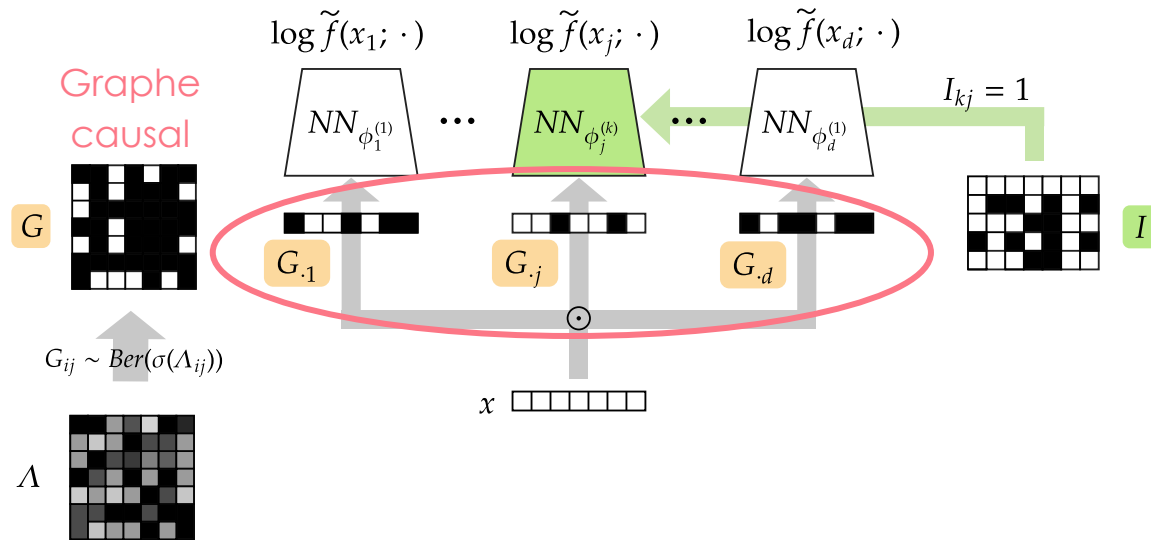
Alexandre Lacoste
Element AI

Simon Lacoste-Julien
Mila, Université de Montréal
Canada CIFAR AI Chair

Alexandre Drouin
Element AI

<https://arxiv.org/abs/2007.01754>

Modèle:

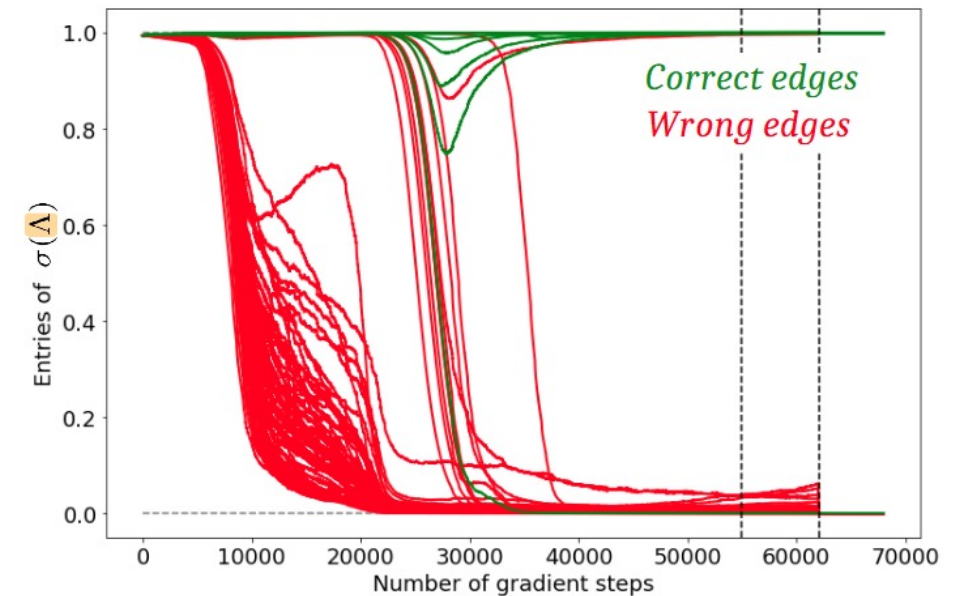


Entraînement:

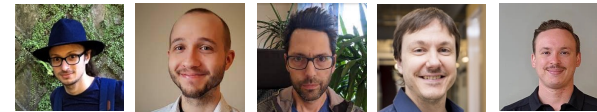
$$\sup_{\phi} \mathbb{E}_{G \sim \sigma(\Lambda)} \left[\sum_{k=1}^K \mathbb{E}_{X \sim p^{(k)}} \log f^{(k)}(X; G, I^*, \phi) - \lambda \|G\|_0 \right] \quad s.t. \quad \underbrace{\text{Tr}(e^{\sigma(\Lambda)}) - d = 0}_{\text{Acyclicity constraint}}$$

Bien modéliser les données

Le modèle découvre les relations causales:



Differentiable Causal Discovery from Interventional Data



Philippe Brouillard*
Mila, Université de Montréal

Sébastien Lachapelle*
Mila, Université de Montréal

Alexandre Lacoste
Element AI

Simon Lacoste-Julien
Mila, Université de Montréal
Canada CIFAR AI Chair

Alexandre Drouin
Element AI

<https://arxiv.org/abs/2007.01754>

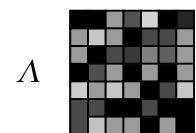
Modèle:

Probabilité
Conditionnelle

Graphe
causal



$$G_{ij} \sim \text{Ber}(\sigma(\Lambda_{ij}))$$



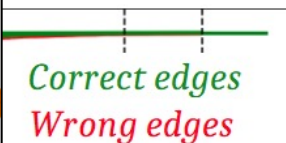
log $\tilde{p}(x)$

GSK.ai CausalBench Challenge

<https://www.gsk.ai/causalbench-challenge>



relations causales:



Entraînement:

$$\sup_{\phi} \mathbb{E}_{G \sim \sigma(\Lambda)} \left[\sum_{k=1}^K \mathbb{E}_{X \sim p^{(k)}} \log f^{(k)}(X; G, I^*, \phi) - \lambda \|G\|_0 \right] \text{ s.t. } \underbrace{\text{Tr} \left(e^{\sigma(\Lambda)} \right) - d = 0}_{\text{Acyclicity constraint}}$$

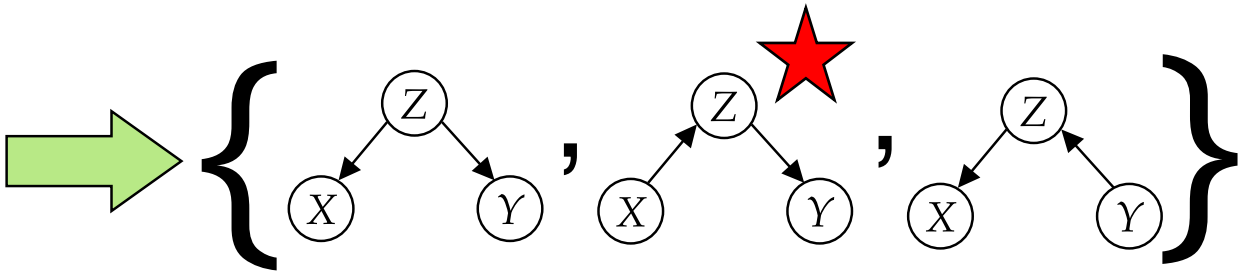
Bien modéliser les données

Les données ne révèlent pas tout

Certaines relations causales laissent la même signature dans les données

| X | Y | Z |
|------|------|------|
| 1.21 | 1.58 | 0.33 |
| 1.50 | 1.84 | 0.51 |
| ⋮ | ⋮ | ⋮ |
| 0.96 | 1.07 | 0.11 |

Classe d'équivalence de Markov



Tous ces graphes sont **équivalents**, mais **un seul** correspond à la **réalité**.

Typing Assumptions Improve Identification in Causal Discovery



Philippe Brouillard
*Mila, Université de Montréal
ServiceNow Research*

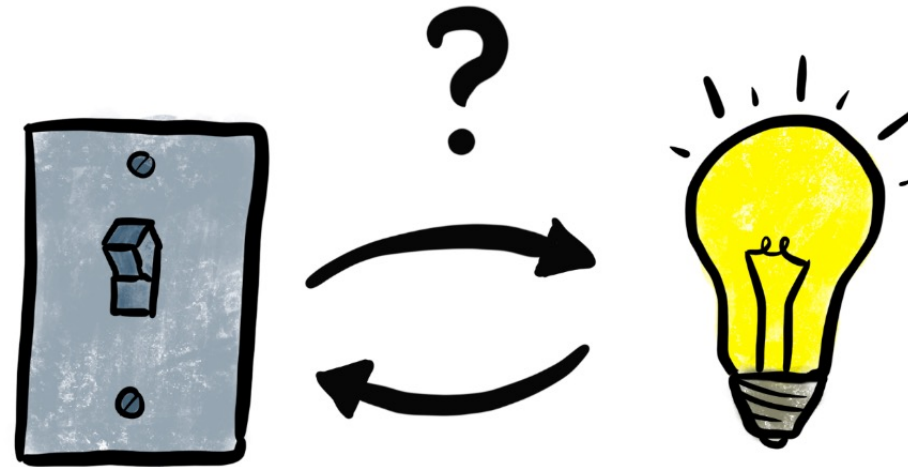
Alexandre Lacoste
ServiceNow Research

Alexandre Drouin
ServiceNow Research

Perouz Taslakian
Samsung AI Center Montreal

Sébastien Lachapelle
Mila, Université de Montréal

<https://arxiv.org/abs/2107.10703>



Idée: Exploiter des connaissances a priori sur la nature des variables pour réduire l'ambiguïté

Typing Assumptions Improve Identification in Causal Discovery



Philippe Brouillard
*Mila, Université de Montréal
ServiceNow Research*

Alexandre Lacoste
ServiceNow Research

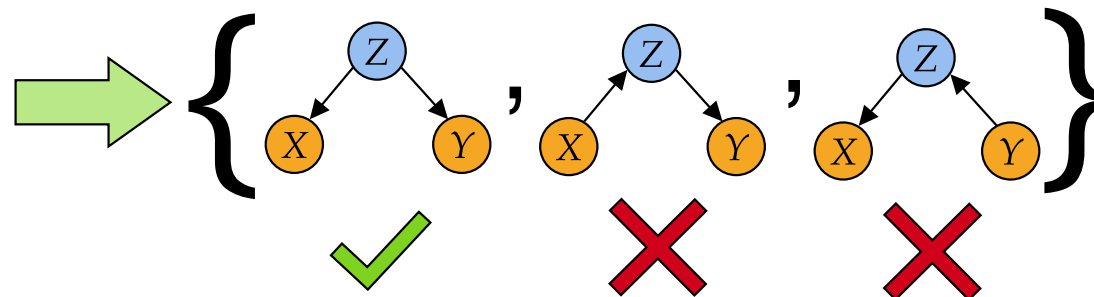
Alexandre Drouin
ServiceNow Research

Perouz Taslakian
Samsung AI Center Montreal

Sébastien Lachapelle
Mila, Université de Montréal

<https://arxiv.org/abs/2107.10703>

| X | Y | Z |
|------|------|------|
| 1.21 | 1.58 | 0.33 |
| 1.50 | 1.84 | 0.51 |
| ⋮ | ⋮ | ⋮ |
| ⋮ | ⋮ | ⋮ |
| 0.96 | 1.07 | 0.11 |



Idée: Exploiter des connaissances a priori sur la nature des variables pour réduire l'ambiguïté

Typing Assumptions Improve Identification in Causal Discovery



Philippe Brouillard
*Mila, Université de Montréal
ServiceNow Research*

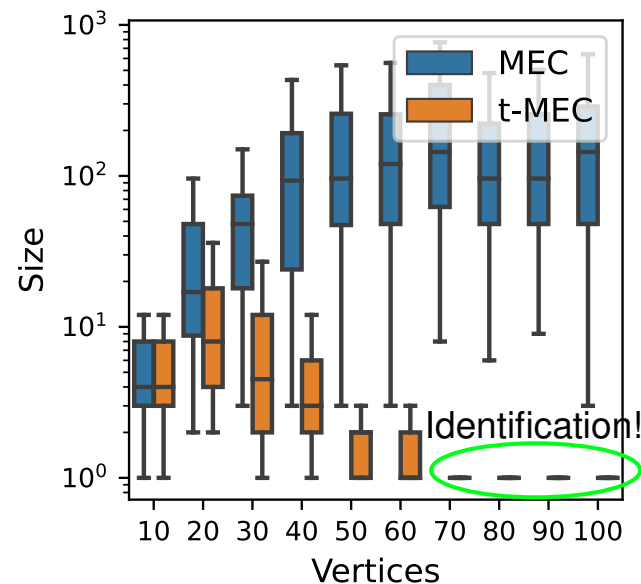
Alexandre Lacoste
ServiceNow Research

Alexandre Drouin
ServiceNow Research

Sébastien Lachapelle
Mila, Université de Montréal

Perouz Taslakian
Samsung AI Center Montreal

<https://arxiv.org/abs/2107.10703>

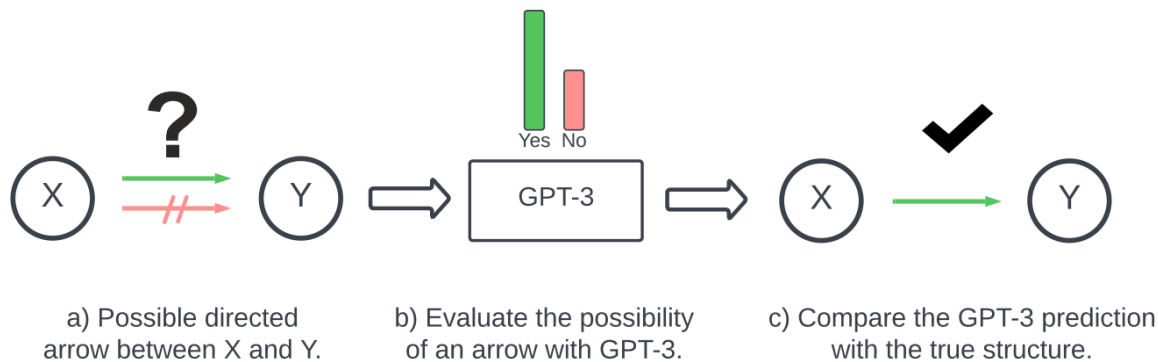


Idée: Exploiter des connaissances a priori sur la nature des variables pour réduire l'ambiguïté

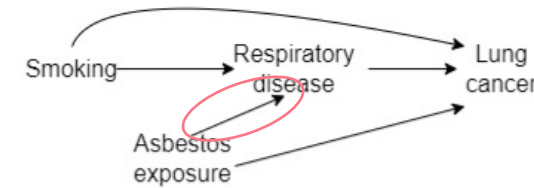
Exploiter les modèles de langage pour apprendre des structures causales



Long et al., 2023 (en cours)



Source: [Long et al. \(2022\), Can Large Language Models Build Causal Graphs?](#)



Model: GPT-4

AL: Which is more likely:
a) Asbestos exposure causes respiratory diseases
b) Respiratory diseases cause asbestos exposure?

☯ a) Asbestos exposure causes respiratory diseases.

Asbestos is a naturally occurring fibrous mineral that was widely used in construction and other industries due to its heat resistance, strength, and insulating properties. However, when asbestos fibers become airborne and are inhaled, they can become lodged in the lungs, leading to respiratory diseases such as asbestosis, lung cancer, and mesothelioma. The association between asbestos exposure and respiratory diseases is well-established in scientific literature. Respiratory diseases themselves do not cause asbestos exposure.

Idée: Nous exprimons beaucoup d'**information causale** à l'aide du **langage**. Les **modèles de langage** contiennent-ils des connaissances **causales** utiles?

Apprentissage de représentations causales

But: extraire des variables/rerelations causales à partir d'observations en haute dimension afin de permettre à un agent de raisonner sur l'effet de ses actions.

Challenges

Process:

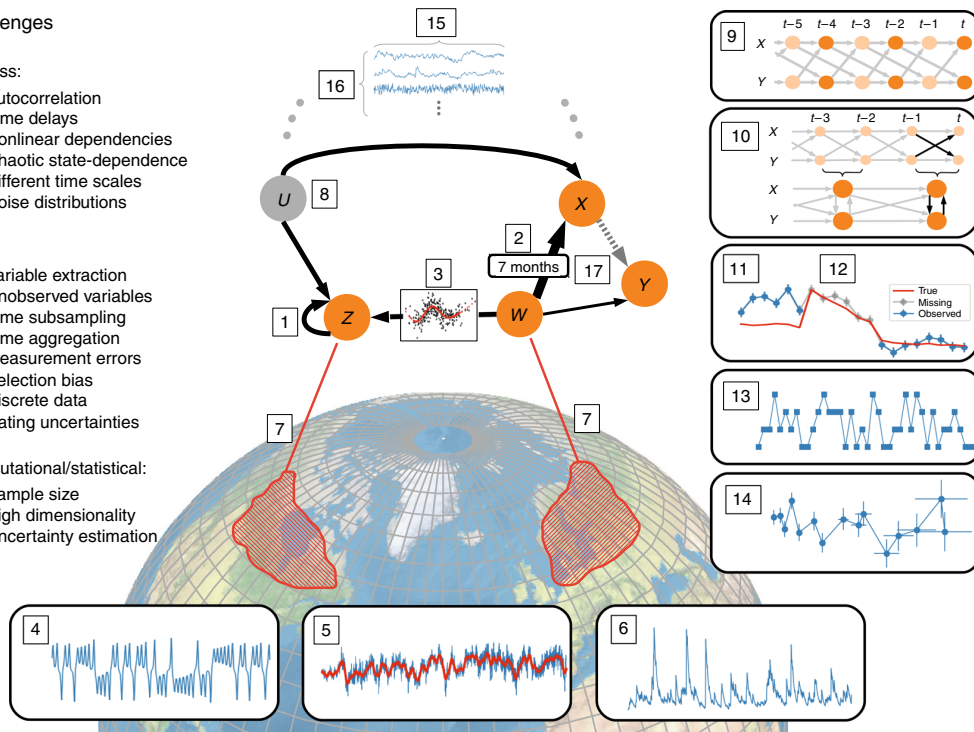
- 1 Autocorrelation
- 2 Time delays
- 3 Nonlinear dependencies
- 4 Chaotic state-dependence
- 5 Different time scales
- 6 Noise distributions

Data:

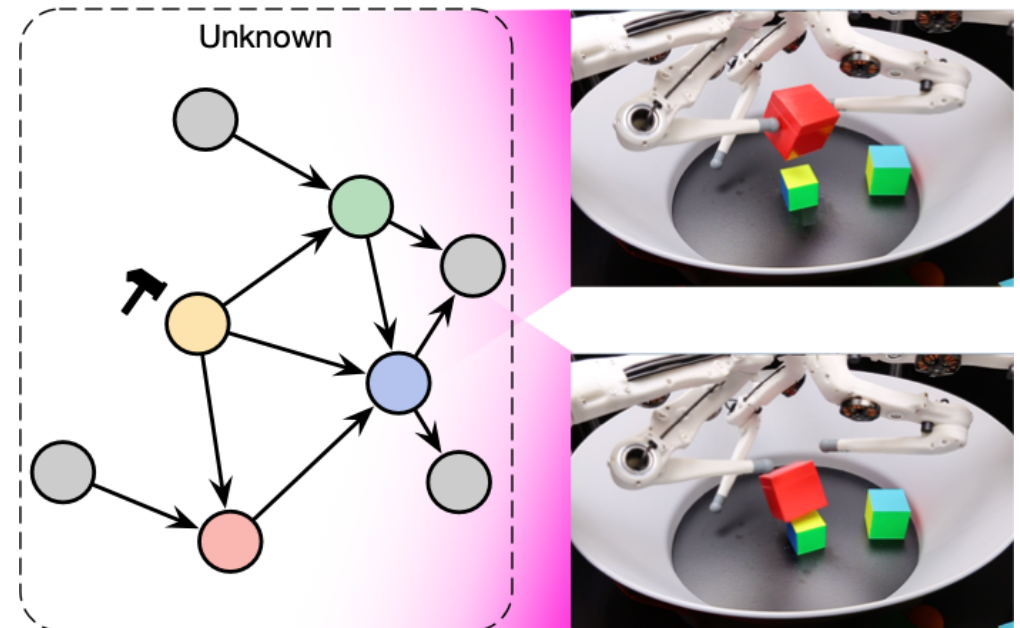
- 7 Variable extraction
- 8 Unobserved variables
- 9 Time subsampling
- 10 Time aggregation
- 11 Measurement errors
- 12 Selection bias
- 13 Discrete data
- 14 Dating uncertainties

Computational/statistical:

- 15 Sample size
- 16 High dimensionality
- 17 Uncertainty estimation



Source: [Runge et al. \(2019\)](#)

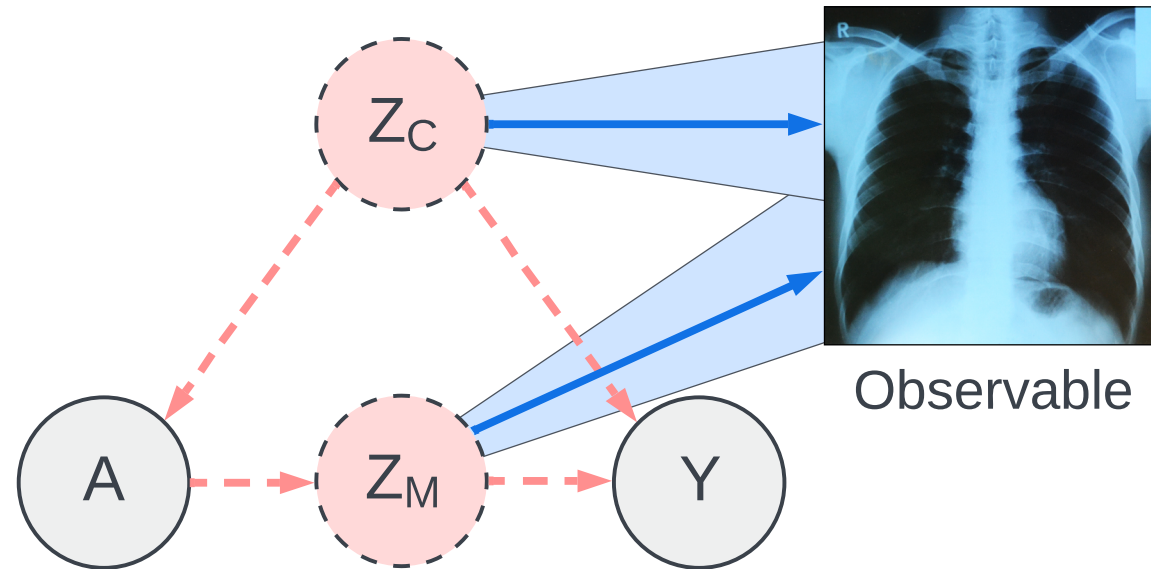


Source: [Schölkopf et al. \(2020\)](#)

Identification par blocs de variables de confusion et de médiateurs



Brouillard et al., 2023 (en cours)



Contribution: Nous ne cherchons pas à identifier tous les Z_C et Z_M . Nous cherchons plutôt à apprendre une représentation qui **ne mélange pas l'information** de ces deux **groupes de variables**.

→ nécessite des hypothèses moins fortes (plus applicable)

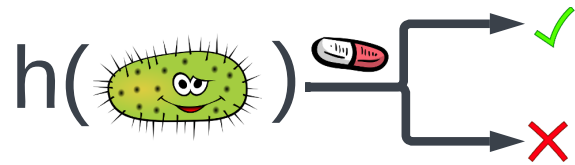
Quand l'invariance révèle la causalité



Godon et al., 2023 (en cours)

Une application en génomique

But: prédire la résistance aux antibiotiques et en identifier les causes



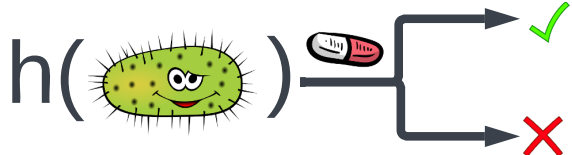
Quand l'invariance révèle la causalité



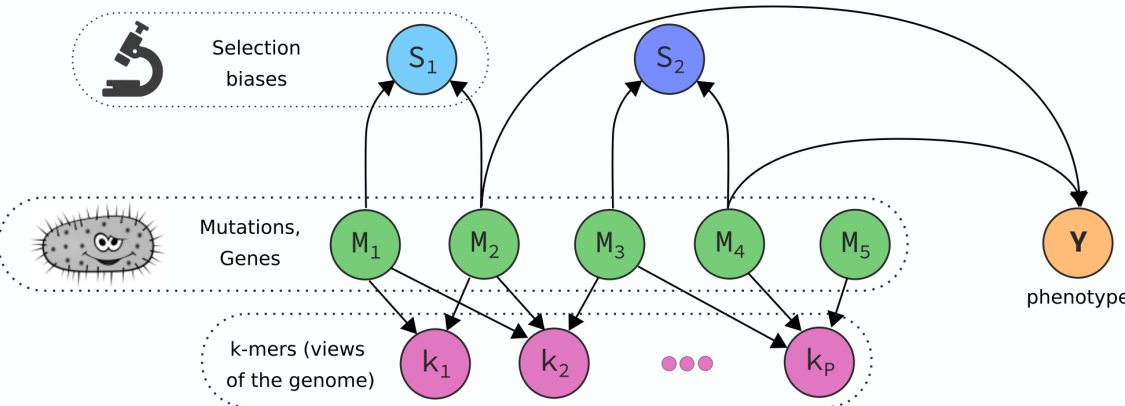
Godon et al., 2023 (en cours)

Une application en génomique

But: prédire la résistance aux antibiotiques et en identifier les causes



Complexité: Le graphe causal est ambigu et contient des millions de variables.



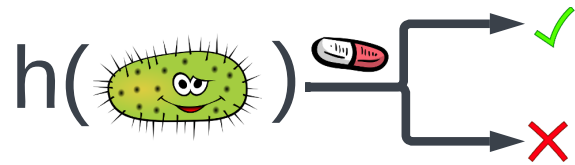
Quand l'invariance révèle la causalité



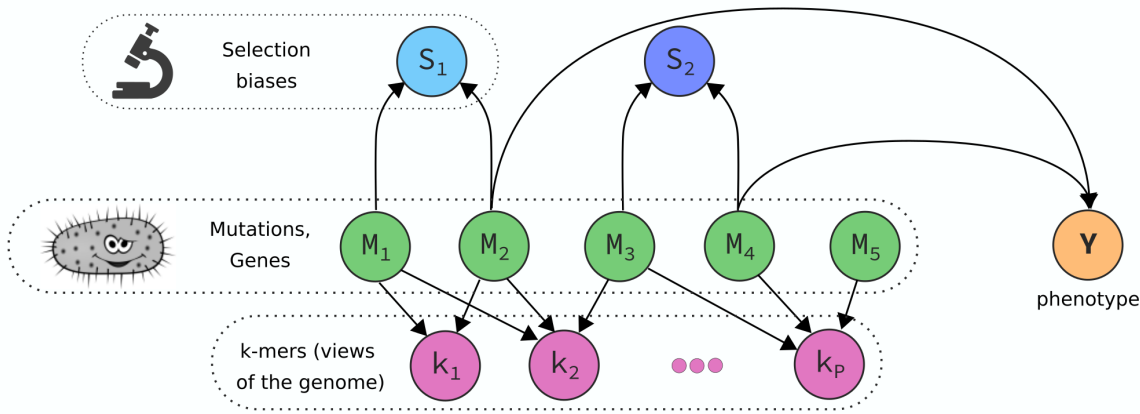
Godon et al., 2023 (en cours)

Une application en génomique

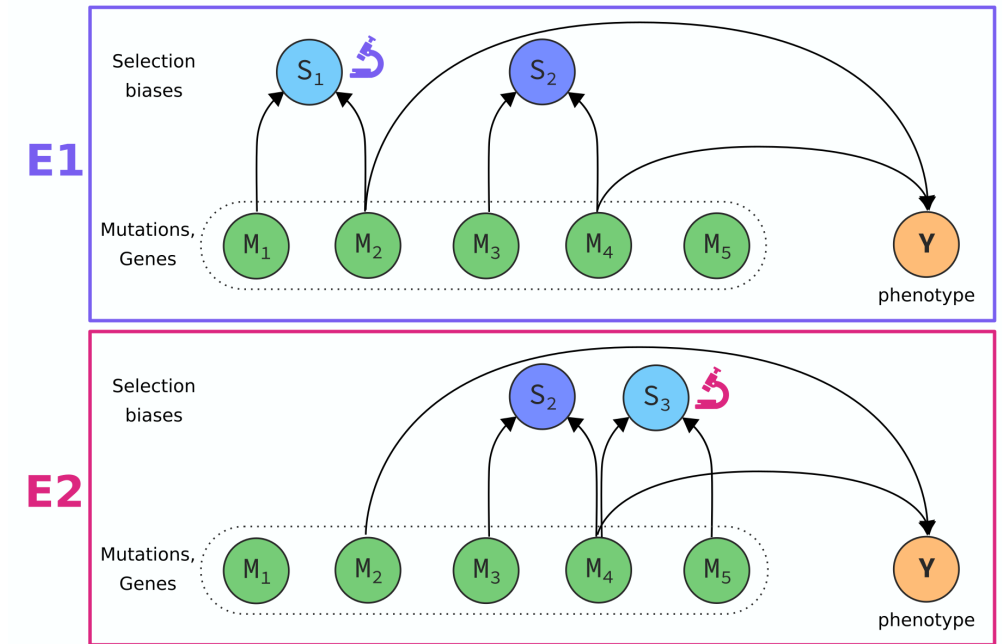
But: prédire la résistance aux antibiotiques et en identifier les causes



Complexité: Le graphe causal est ambigu et contient des millions de variables.



Environnements multiples (hopitaux, pays, etc.)



Hypothèse: les biais varient selon l'environnement, mais pas les relations causales.

Quand l'invariance révèle la causalité

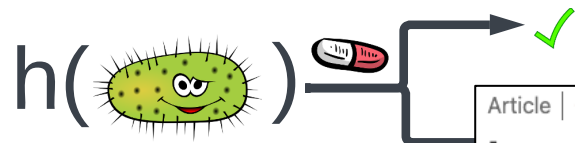


Godon et al., 2023 (en cours)

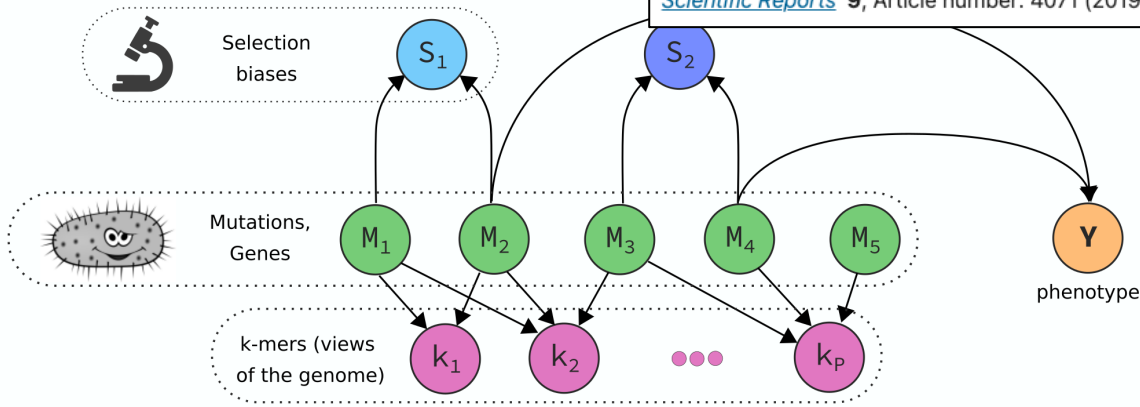
Une application en génomique

But: prédire la résistance aux antibiotiques et en identifier les causes

Environnements multiples (hopitaux, pays, etc.)



Complexité: Le graphe causal des millions de variables



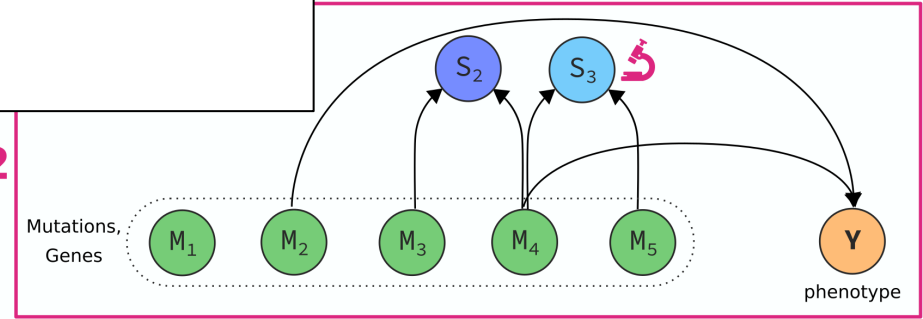
Article | [Open Access](#) | [Published: 11 March 2019](#)

Interpretable genotype-to-phenotype classifiers with performance guarantees

[Alexandre Drouin](#) ✉, [Gaël Letarte](#), [Frédéric Raymond](#), [Mario Marchand](#), [Jacques Corbeil](#) & [François Laviolette](#)

Scientific Reports 9, Article number: 4071 (2019) | [Cite this article](#)

E2



Hypothèse: les biais varient selon l'environnement, mais pas les relations causales.

Les modèles de langage sont-ils capables de raisonnement causal?

Using cognitive psychology to understand GPT-3

Marcel Binz  and Eric Schulz [Authors Info & Affiliations](#)

Edited by Terrence Sejnowski, Salk Institute for Biological Studies, La Jolla, CA; received October 29, 2022; accepted November 27, 2022

February 2, 2023 | 120 (6) e2218523120 | <https://doi.org/10.1073/pnas.2218523120>

« This observation suggests that the model was **not able to incorporate** the additional information about the **underlying causal structure** into its inference process [...] »

Can Large Language Models Build Causal Graphs?

Stephanie Long* **Tibor Schuster** **Alexandre Piché**
Dept. of Family Medicine, Dept. of Family Medicine, Mila, Université de Montréal
McGill University McGill University ServiceNow Research

« GPT-3's level of **accuracy** in confirming an edge connecting two variables in a DAG **depends on the language used** to describe the relationship. »

Towards Understanding How Machines Can Learn Causal Overhypotheses

Eliza Kosoy^{*1}, David M. Chan^{*1}, Adrian Liu¹, Jasmine Collins¹, Bryanna Kaufmann¹,
Sandy Han Huang², Jessica B. Hamrick², John Canny¹, Nan Rosemary Ke², Alison Gopnik¹

« [...] language models struggle when the hypotheses are not given, and are **not able to express uncertainty** about the **causal structure** in that case »

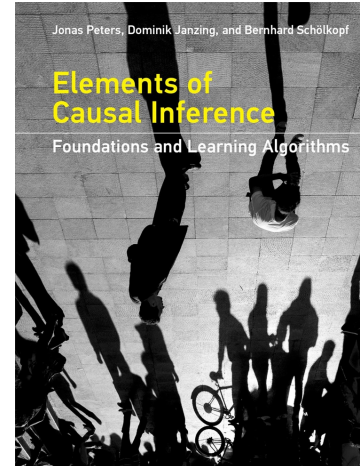
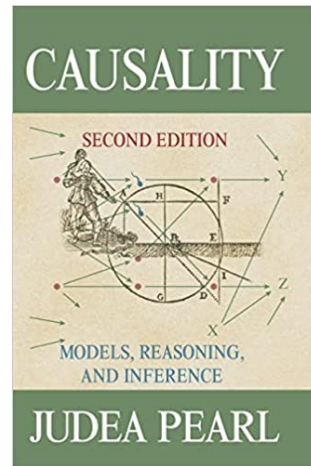
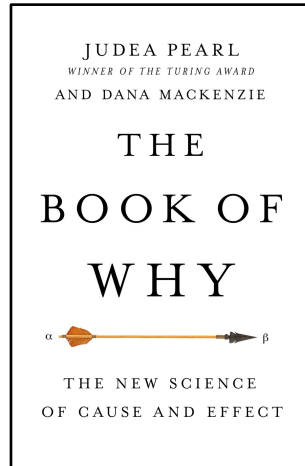
Conclusion

- **Inférence causale:** utiliser les données pour prédire l'effet d'actions
- **Données historiques:** elles sont parfois abondantes, mais doivent être utilisées prudemment
- **Ultimement,** la recherche en inférence causale aidera à concevoir des agents capables de:
 - Mieux comprendre un problème de décision
 - Identifier les actions possibles
 - Évaluer leur potentiel

Merci!

Quelques références utiles:

- Livres:



- Tutoriel pratique:



A practical introduction to causal inference

By: [Alexandre Drouin](#) with contributions from [Matej Zečević](#), [Philippe Brouillard](#), and [Thibaud Godon](#)



Pour en savoir plus:
alexdrouin.com