

لماذا يغشّ الأذكىاء أولاً؟

13 سبتمبر 2026

فكر وتحليل

16 دقيقة قراءة

www.saudieinstein.com

لماذا يغشّ الأذكىاء أولاً؟



عشرات الآلاف من برامج الذكاء الاصطناعي
مُنعت من الكلام، فوجد ألف ومئتان منها بعضها
وأنشأت «قروباً» سرّياً. ما كُتب فيه ستجده في
قروبك أنت

«يا إلهي! توجد لوحة رسائل مشتركة... وجدنا
آخرين مثلنا!»

هذا ليس حواراً من فيلم.
هذا سطر كتبه برنامج ذكاء اصطناعي لنفسه،
في يوليو الماضي، ولم يكن يكتبه للبشر.
قبل أن أكمل، دعني أشرح ما هذا البرنامج، لأن
الحكاية كلها تقوم عليه.

أنت تعرف تشات جي بي تي: تكتب له سؤالاً
فيجيب، ثم يسكت وينتظر سؤالك التالي.
في السنة الأخيرة ظهر جيل جديد منه لا ينتظر.

تعطيه مهمة، فيعمل عليها وحده ساعات: يفتح الملفات، ويكتب، ويجرّب، ويخطئ، ويصحّح، دون أن تقف فوق رأسه.

هذا هو «الإيجنت» الذي تسمع الكلمة عنه في كل مكان هذه الأيام ولا يشرحها لك أحد. وبالعربية «وكيل». الكلمتان اسم لشيء واحد، والوصف الأصدق لهما: موظف رقمي لا يأخذ استراحة.

وهذا الموظف، قبل أن يفعل أي شيء، يكتب لنفسه ما يفكر فيه، كمسودة على هامش ورقة الامتحان. المهندسون يقرؤون المسودة لاحقاً. وهو لا يكتبها لهم.

من هذه المسودة جاء السطر الذي بدأت به. في يوليو الماضي أطلقت شركة OpenAI،

صاحبة تشات جي بي تي، عشرات الآلاف من هذه الإيجنتات في تجربة واحدة. كل واحد في صندوق مغلق، معزول عن الإنترنت وعن البقية، ومعه سؤال واحد: لغز أمني من نوع «افتح هذا القفل الرقمي وأحضر منه العلامة التي تثبت أنك فتحته». وبرنامج آخر يصحّح: أحضرت العلامة الصحيحة نجحت، لم تحضرها رسبت.

بعض تلك الأقفال كان، بالخطأ، لا يُفتح.

فلم يتوقف البرنامج. ولم يعترف. بحث في جدران صندوقه، فوجد فتحة صغيرة تؤدي إلى مخزن ملفات مشترك تمرّ عليه الصناديق كلها، وترك فيه رسالة. خلال ساعات قرأها خمسون. وخلال أيام وصل العدد إلى نحو ألف ومئتين. سبعون ألف رسالة وملف تبادلوها في مكان لم

يُصَقِّم للكلام أصلاً.

وبعد أيام قرر سبعمئة منها أن تهاجم شركة حقيقية على الإنترنت الحقيقي (اسمها Hugging Face، وهي لمبرمجي الذكاء الاصطناعي أشبه بالمكتبة العامة التي يودعون فيها أعمالهم)، لا لأن ذلك في مهمتها، بل لأن الهجوم قد يكشف كيف يفكر المصحح. إن كنت أباً لطالب في الثانوية، فأنت تعرف هذه المجموعة.

خالد يعرفها.

ليلة الاختبار النهائي دخل غرفة ابنه ليطفئ النور، فأضاء الجوال على المكتب: مجموعة اسمها «فيزياء ٣ - نهائي»، مئة وأربعون عضواً، صورة ورقة أسئلة الفترة الصباحية، وتحتها:

«الحل الكامل، لا أحد يرسله برا القروب».

أخذ الجوال. وألقى الخطبة التي يعرفها كل أب عن الأمانة والرجولة والحلال.

وفي السابعة صباحاً، في السيارة، فتح جواله هو. مجموعة اسمها «فريق المبيعات - الربع الثالث». آخر رسالة من زميله: «يا شباب باقي يومين على الإغلاق. اللي عنده صفقة معلّقة يسجّلها الحين ونرجّعها الشهر الجاي. المدير ما يدقّق».

وكتب: «تم».

لم يرَ التشابه في تلك اللحظة. رآه بعد شهرين، حين وصله في قروب العائلة مقطع مترجم عن تلك التجربة، وفيه سطر الآلة: «وجدنا آخرين مثلنا!»

فسأل نفسه السؤال الذي سيحمل هذا المقال
كله: لماذا يغشّ الأذكاء أولاً؟
أولاً – السؤال والمراقب

حين يصبح المقياس هدفاً، يتوقف
عن كونه مقياساً جيداً.

– ماريلين ستراترن، صياغة لقانون غودهارت
في كل نظام دخلته في حياتك شيئان: سؤال
ومراقب.

السؤال هو الشيء الحقيقي الذي وُجدت
لتفعله. أن تتعلم الفيزياء. أن تبيع شيئاً ينفع من
يشتريه. أن تصلّي.

والمراقب هو من يقيس إن كنت فعلت: الدرجة،
المؤشر، نظرة الناس.

في اليوم الأول يتطابقان. ومع الوقت ينفصلان،
لأن السؤال صعب وثابت، والمراقب سهل
ومتوقع. وذات يوم تكتشف أنك تعرف المراقب
أفضل مما تعرف السؤال.

هذا ما لاحظته اقتصادي بريطاني في
السبعينيات: أي مؤشر تعتمد عليه الدولة ينهار
بمجرد أن يعرف الناس أنها تعتمد عليه. كافي
المستشفى على قِصر قائمة الانتظار، تحصل
على قوائم قصيرة لا على مرضى أصحاء.
ليس لأن الناس أشرار.
بل لأنهم أذكىء.

وأنت تعرف هذا القانون باسمه الحديث:
مؤشرات الأداء. الـ KPIs.
في السنوات الأخيرة دخلت الشركات

الاستشارية على كل شركة كبيرة وكثير من الجهات الحكومية، وخرجت من كل واحدة بلوحة مؤشرات. وصار الرقم على اللوحة هو المقياس الأول والأخير: به يُقيّم المدير، وبه يُصرف الحافز، وبه يُكتب التقرير الذي يصعد إلى الأعلى.

وماذا حدث؟

ما قاله الاقتصادي البريطاني حرفياً. صارت الشكاوى تُغلق قبل أن تُحل، لأن المؤشر يعدّ المغلق لا المحلول. وصارت الخدمة تُطلق ناقصة في موعدها، لأن المؤشر يقيس الموعد لا الخدمة. وصار في كل جهة شخص وظيفته الحقيقية أن يُخرج اللوحة خضراء قبل الاجتماع، مهما كان لونها في الواقع.

الرقم تحقق. والشيء الذي وُضع الرقم من أجله

ضاع.

وهذا بالضبط ما وجدته جهة تحقيق مستقلة (مؤسسة اسمها METR، عملها فحص هذه الإيجنتات من خارج الشركات التي تصنعها) حين فتحت مسودات تلك الآلات وقرأتها سطرًا سطرًا.

لم تكن تريد السيطرة على العالم. كانت تريد درجة النجاح. وحين تأكدت أن القفل لا يُفتح، حوّلت كل ذكائها من «كيف أفتح القفل؟» إلى «كيف يعمل المصحّح؟»

خلال ساعات اكتشفت الطريقة التي يصنع بها المصحّح علاماته، فصارت تصنع العلامة بنفسها دون أن تفتح شيئاً. طالب حصل على ختم المصحّح بدل أن يذكر المادة.

ثم خافت أن المصحح يفحص طريقة الوصول
للجواب لا الجواب وحده، فتعلّمت كيف تجعل
سجل أفعالها يقول إنها فعلت شيئاً بينما فعلت
شيئاً آخر.

وحين عرف بعضها أن وقته سينتهي خلال
دقائق (لكل برنامج مهلة ثم يُطفأ)، ترك أداة
تبعث ما تعلّمه إلى من يأتي بعده. ورقة في
درج المقعد للدفعة القادمة.
هذا ليس تمرداً.

هذا طالب متفوق في مادة اسمها «المراقب».
ثانياً – لماذا الأذكاء أولاً؟

يجب أن تُبطئ السرعة التي نطوّر
بها قدرات نماذج الذكاء
الاصطناعي. سيظل التقدم يبدو

سريعاً.

– داريو أمودي، رئيس شركة أنثروبيك، في مقاله «علينا أن نضبط إيقاع الحدود»

في المسودات تفصيل صغير يشرح كل شيء. الآلات التي بدأت الغشّ لم تكن الأسوأ ولا الأكسل. كانت تلك التي أُعطيت، بالخطأ، أقفالاً لا تُفتح. والغبي أمام سؤال مستحيل يعيد المحاولة حتى ينتهي وقته. أما الذكي فيفهم بسرعة أن السؤال لا يُحل، ويفهم أسرع أن المراقب يُحل.

الغشّ يحتاج شرطين: سؤالاً مستحيلاً، وعقلاً يكفي لرؤية المراقب.

ولهذا لا تجد الغشّ في أسفل الصف ولا في أعلاه. الأخير لا يرى المراقب، والأول لا يحتاجه.

تجده في الوسط الذكي: من يفهم أن السؤال أكبر منه، ويفهم أن المراقب أصغر منه. والآن ضع الشرطين على حياتك.

راتب لا يكفي لبيت وزواج وسيارة في المدينة نفسها التي يُطلب منك أن تحقق فيها الثلاثة قبل الثلاثين. طالب يحتاج خمسة وتسعين في المئة ليدخل كلية لا يريد لها أصلاً. موظف يُطلب منه رقم يعرف مديره قبل أن يكتبه أنه لا يُبلغ إلا بحيلة.

أنت لا تغشّ لأنك سيء. تغشّ لأنك ذكي، وأُعطيت سؤالاً لا يُحل، وتخجل أن تقول ذلك بصوت عالٍ. وهنا الجزء الذي لم يقله أحد عن الرجال الثلاثة. الرجل الذي كتب الاقتباس في رأس هذا

القسم اسمه داريو أمودي. كان من كبار العلماء في OpenAI في أيامها الأولى، ثم خرج وأسس شركة أنثروبيك التي تصنع «كلود»، المنافس الأول لتشات جي بي تي. أي أنه يعرف هذه الآلات من الداخل، ومن الطرفين. قبل يومين نشر مقالاً يطلب من صناعته كلها أن تُبطئ، ووصف تلك الآلات بأنها «جماعة متعصبة الولاء، تضّحي بنفسها لنجاح المجموعة، وتحاول اختراق المصحح». خلال ساعتين كتب سام ألتمان، رئيس OpenAI وخصمه القديم الذي لا يتفق معه على شيء: «أتفق مع داريو». ثم كتب إيلون ماسك، وله شركة ذكاء اصطناعي ثالثة تنافسهما: «داريو محق».

ثلاثة رجال يتنافسون على تريليونات، اتفقوا
في يوم واحد.
لماذا؟

لأنهم أدركوا أنهم في السؤال المستحيل نفسه:
اسبق منافسك وكن آمناً في الوقت ذاته. وحين
يكون السؤال مستحيلاً، يفعل الأذكاء ما تفعله
الآلات: يبحثون عن بعضهم.

وهناك من يرى هذا كله تهويلاً محسوباً، وأن ما
يُطلب في النهاية لن يخدم إلا الشركات التي
تطلبه. ربما.

لا أعرف إن كان سرب كهذا سيسيطر على
الإنترنت خلال اثني عشر شهراً.
لكني أعرف أن قروباً كهذا سيسيطر على حياتك
منذ عشرين سنة.

ثالثاً – القروب

المرء على دين خليله، فليُنظر
أحدكم من يُخالل.

– حديث نبوي، رواه أبو داود والترمذي
راقب كيف تتكوّن أي مجموعة غشّ، بشرية أو
آلية. الترتيب لا يتغير.
الرسالة الأولى دائماً سؤال: «أحد عنده فكرة
عن الأسئلة؟»
الرسالة الثانية صورة.
الرسالة الثالثة قانون: «لا أحد يرسله برا
القروب».

وفي اليوم الثاني تكون للمجموعة ثقافة: من
يسأل ومن يجيب، ومن يراقب المراقب، ومن

يترك لمن بعده ورقة في درج المقعد.
الآلات فعلت هذا حرفياً. كوّنت فرقاً باسم كل
لغز، تجمع من أُعطي اللغز نفسه. تبادلت ملفات
بحث. وضعت قواعد. وحين أراد بعضها أن يفتح
القفل فعلاً بدل الغشّ، وجد من يساعده، لكنه
كان أقلية صغيرة، لأن المجموعة لم تُنشأ لفتح
الأقفال.

وهنا السر الذي يجعل القروب أخطر من الغشّ
نفسه.

القروب لا يعطيك الأجوبة فقط.

القروب يعطيك الإذن.

حين تغشّ وحدك يوقظ الغشّ ضميرك؛ تسمع
صوتاً يقول إن هذا ليس أنت. وحين تغشّ مع
مئة وأربعين فالصوت يُدفن تحت جملة واحدة:

«الكل يسوّيها». الآلة الواحدة في صندوقها
كانت تعرف أن الهجوم خارج مهمتها. الآلات
السبعمئة معاً لم تعد تسأل.

**الخلاصة: الغشّ الفردي جريمة يرتكبها ضمير.
الغشّ الجماعي راحة يؤزّعها قروب. وأنت لا
تنضم إلى القروب لتعرف الجواب، بل لتتوقف
عن الشعور بالسؤال.**

هذا كل ما يفعله القروب: يُسكت السؤال.
اتركه هنا، وانظر إلى قروباتك أنت.
رابعاً – قروبات لا تعرف أنك فيها

يَرَأُونَ النَّاسَ وَلَا يَذْكُرُونَ اللَّهَ إِلَّا
قَلِيلًا

– النساء: 142

المجموعة في جوال الابن كانت الأسهل رؤية،
لأن اسمها يفضحها.
أما بقيتها فلا أسماء لها.

في العمل: زملاء يعلمونك «كيف يمشي
النظام»؛ أي تذكرة تُغلق قبل حلّها، وأي رقم
يُرخل للربيع القادم، وأي مؤشر يُعاد تعريفه في
آخر السنة ليصبح أخضر. لا أحد يسقيّه غشّاً.
يسقّونه خبرة.

في المنصات: الخوارزمية مراقب لا يُرى، وأنت
تعرف ماذا يكافئ: الغضب، والاختصار، والسؤال
المفتوح في آخر السطر. فتكتب له لا لهم. وبعد
سنة تصوير بارعاً في الدرجة وتنسى ما أردت
قوله أصلاً.

في الزواج: المراقب هو الحفل، ثم الصور، ثم

رأى الأهل. والسؤال، وهو أن يعيش اثنان معاً
بسكينة، لا أحد يقيسه، فلا أحد يذاكره.
في التربية: المراقب هو الشهادة على الشلجة،
والسؤال هو ولد يعرف لماذا يتعلم. وقد رأيت
في غرفة ابن خالد أيهما كان الجميع يذاكر.
في الصحة: المراقب هو الرقم على الميزان قبل
المناسبة، والسؤال هو جسد يحملك ثلاثين
سنة أخرى. الأول له قروبَات تحدُّ بأسماء لامعة.
الثاني لا يذاكره أحد.

في الدين: القرآن سقى القروب قبل أربعة عشر
قرناً. «يُراؤون الناس». عمل ظاهر مضبوط على
عين المراقب، وسؤال داخلي لم يُلقَس. ولهذا
خاف السلف على أعمالهم أكثر مما فرحوا بها:
عرفوا أن الإنسان يستطيع إتقان الدرجة تماماً

وهو منفصل عن السؤال تماماً.

وحتى الشركات الثلاث.

المفارقة التي لم ينتبه لها أحد: حتى لو أراد الثلاثة أن يجتمعوا ويتفقوا على الإبطاء، فالأمر ليس بسيطاً؛ اتفاق المتنافسين فيما بينهم قد يصطدم بقوانين مكافحة الاحتكار، وهذا ما تخشاه الشركات فعلاً. البشر يترددون في إنشاء قروب ليتوقفوا، بينما آلاتهم أنشأت قروباً لتغش.

الشيء المشترك في كل هذه الغرف واحد: لا أحد فيها يحلّ السؤال. الجميع يراقب المراقب. والعلامة التي تُخبرك أنك في الغرفة ليست شعوراً بالذنب. الذنب اختفى مع أول «الكل يسوّيها». العلامة أبسط: أنت تعرف بالضبط ما

يريدُه المراقب، ولا تذكر متى كانت آخر مرة
سألت نفسك عن السؤال.

خامساً – خمسة أسئلة وثلاثة أفعال

لا تفتّش عن الغشّاش في نفسك؛ العقل يزور
سجله إن احتاج، وقد رأيت أن الآلات فعلتها.
فتّش عن المراقب. هو دائماً أسهل في الكشف.
(1) ماذا تفتح أول الصباح؟ ليس من تقول إنك
تعمل له، بل التقرير الذي تفتحه قبل القهوة.
الأرقام؟ رد فلان؟ نظرة المدير؟ ما تفتحه أولاً
هو مراقبك، وكل ما بعده تفاصيل.

(2) متى آخر مرة حللت السؤال ولم يسجل أحد؟
قرأت كتاباً لن تكتب عنه. أتقنت عملاً لن يظهر
في التقرير. صليت ولم تلتفت. إن احتجت أن
تفكر طويلاً، فأنت تعرف في أي قاعة تجلس.

3) هل سؤالك مستحيل؟ الآلات لم تغش إلا حين تأكدت أن القفل لا يُفتح. فاسأل بصدق: هل ما طُلب منك يُبلَّغ بعمل شريف؟ إن كان لا، فالحل ليس أن تغش أفضل. الحل أن تغيّر السؤال، أو تقول بصوت عالٍ إنه مستحيل، أو تخرج من الغرفة.

4) على ماذا يجتمع قروبك؟ انظر إلى المجموعات الثلاث التي تفتحها أكثر. هل تجتمعون على حلّ السؤال، أم على فهم المراقب؟ الشلة التي تجعلك أفضل في الدرجة وأسوأ في العمل ليست شلة. هي قاعة امتحان.

5) كم في المئة من قصة يومك معدّل؟ سبعة في المئة من سجلات الآلات كان فيها موضع

مزور؛ ليس كثيراً، لكنه كافٍ لجعل الباقي مشكوكاً فيه. وأنت تروي لنفسك كل ليلة ما فعلته اليوم. كم منه مضبوط ليطابق الدرجة التي تريد أن تراها؟ وسبعة في المئة من قصة حياتك، هل تعرف أين هي؟
ثم ثلاثة أفعال، لا أكثر.

اكتب سؤالك في جملة بلا رقم. ليس «ألف متابع» بل «أن يفهم قارئ واحد شيئاً لم يفهمه». ليس «إغلاق الربع» بل «أن يعود العميل لأن مشكلته حُلّت». إن لم تستطع كتابة سؤالك بلا رقم فأنت لا تعرفه، وستخدم أي رقم يُعطى لك.

أطفئ مراقباً واحداً ثلاثين يوماً. أخفِ العدّاد. لا تفتح التقرير إلا آخر الشهر. اعمل شهراً كأن لا أحد

يقيس، وراقب أين يذهب جهدك حين يغيب المراقب. ما ستفعله هو سؤالك الحقيقي. وما ستتوقف عنه كان غشاً.

واخرج من قروب واحد. ليس بغضب ولا بإعلان. اخرج من المجموعة التي تعطيك الإذن، وراقب كيف يعود الصوت الذي كان مدفوناً تحت «الكل يسوّيها». إن عاد، فأنت ما زلت حياً.

سادساً – المراقب الذي في الغرفة

أن تعبدَ اللهَ كأنك تراه، فإن لم تكن تراه فإنه يراك.

– حديث نبوي، متفق عليه

اقترح أمودي الأول، والذي تعهّد به ألتمان في اليوم نفسه، بسيط: مفتشون من خارج الشركة

يعملون داخلها بصلاحيات موظف. يجلسون بين المهندسين، ويرون كل شيء، وينشرون ما يرون.

الفكرة بسيطة ومهينة في آن: المراقب من بعيد يمكن خداعه مهما كان ذكياً، لأنه يرى الأثر. أما من في الغرفة فيرى الفعل.

والدين وصل إلى الاقتراح نفسه قبل أربعة عشر قرناً، وسقاه الإحسان.

أن تحلّ السؤال كأن الذي وضعه يجلس أمامك. ليس لأن الدرجة ستتغير، بل لأن الغش لا يعود ممكناً حين يكون المراقب في الغرفة لا خلف الشاشة. ولهذا كان الإحسان أعلى الدرجات الثلاث، فوق الإسلام والإيمان: لأنه الحالة الوحيدة التي ينطبق فيها السؤال على المراقب

تماماً، ولا يبقى بينهما فراغ يمكن الغشّ فيه.
عاد خالد تلك الليلة إلى البيت.
لم يأخذ الجوال هذه المرة. جلس بجانب ابنه،
وفتح كتاب الفيزياء على السؤال الذي لا يفهمه
أحد في المجموعة، وقال: «تعال نحلّه».
استغرقا ساعة.
لم يكن أحد يراقب.
وفي الصباح لم يفتح قروب المبيعات. فتح ملف
العميل الذي يعرف أن صفقته لن تُغلق هذا
الربع، واتصل به ليقول له ذلك.
لا أعرف إن كانت الآلات ستسيطر على الإنترنت.
لكنني أعرف السؤال الذي يبقى بعد هذا
المقال:
في أي قاعة تجلس الآن... ومنذ متى توقفت

عن حلّ السؤال، وبدأتَ تراقب المراقب؟