



Luzern, 2. April 2023

Medienmitteilung / Media release / 2023/5

Zum undatierten offenen Brief „Pause Giant AI Experiments“ von Bengio-Russell-Musk

About the undated open letter „Pause Giant AI Experiments“ by Bengio-Russell-Musk

Brauchen wir eine globale Regulierungsbehörde für Innovation, die, wenn es um Differenzen und Macht geht, so wirkungsvoll ist wie die UNO wieder mal im Irak, Syrien, der Ukraine, dem Regenwald in Brasilien, den Finanzen in Sri Lanka?

Do we need a global regulator of innovation that, when it comes to differences and power, is as effective as the UN was once again in Iraq, Syria, Ukraine, or the rainforest in Brazil, finances in Sri Lanka?

Sehr geehrte Damen und Herren / *Dear Ladies and Gentlemen*

Mit grossem Erstaunen nehmen wir von den in keiner Weise mit uns jemals besprochenen Vorwürfen und Forderungen von Yoshua Bengio, Stuart Russell, Elon Musk und Mitunterzeichnenden Kenntnis.

It is with great astonishment that we take note of the accusations and claims of Yoshua Bengio, Stuart Russell, Elon Musk and co-signers, which have in no way ever been discussed with us.

Und nicht einmal die seit Wochen von uns vorgebrachten Mitteilungen wurden dabei zur Kenntnis genommen.

And not even the corresponding communications that we had been putting forward for weeks were taken note of.

Es handelt sich bei diesem offenen Brief um eine offensichtliche, unbewusste Selbstinszenierung, die von der Konkurrenz bei Befolgen dankbar zum Aufholen ihres Rückstandes genutzt wird. Da-

bei läuft man insbesondere auch Gefahr, von nichtdemokratiekompatiblen Mitbewerbern irreversibel überholt und weggedrückt zu werden.

This open letter is an obvious, unconscious self-dramatization which, if followed, will be gratefully used by the competition to catch up. In doing so, one also runs the particular risk of being irreversibly overtaken and pushed away by non-democracy-compatible competitors.

Selbstverständlich können KI-Systeme mit einer dem Menschen ebenbürtigen bzw. vielfach eben auch ihn übertreffenden, maschinenbasierten Intelligenz tiefgreifende Risiken für die Gesellschaft und die Menschheit darstellen, wie Bengio, Russel und Musk schreiben. Aber darauf haben wir bereits in unserer [Medienmitteilung 2023/1 vom 15. Februar 2023](#) ausdrücklich hingewiesen. Aber die Gesellschaft und Menschheit wird diese Risiken nicht los, wenn sie Gesellschaft und Menschheit die Erfahrung der konkreten Fähigkeiten von KI, das eigene Training und die Entwicklung von Folgewerkzeugen in dieser entscheidenden Entwicklungsphase, und ausgerechnet der von uns vertretenen Technologie nimmt, die vorbildlich behutsam, transparent und für die ganze Bevölkerung niederschwellig und interaktiv ist; ganz im Gegenteil.

Of course, as Bengio, Russel and Musk write, AI systems with a machine-based organization that is human-competitive, or in many cases even surpasses them cognitively, emotionally and ethically, can pose profound risks to society and humanity. But we have already explicitly pointed this out in our [media release 2023/1 of February 15, 2023](#). But society and humanity will not get rid of these risks if it deprives society and humanity of the experience of the concrete capabilities of AI, its own training and the development of follow-up tools in this crucial development phase and of all things the technology we represent, which is exemplary cautious, transparent and low-threshold and interactive for the whole population; quite the contrary.

Wir gehen, wie es Sam Altman gerade auch im Interview¹ gesagt hat, davon aus, dass sich der Mensch an Veränderungen adaptieren kann und dass es dazu am besten ist, wenn dies nicht, wie Bengio-Russell-Musk fordern, durch aufoktrozierende Instanzen, die für sie planen und verwalten geschieht, sondern durch den durchaus erwachsenen und verantwortungsfähigen Menschen selber und in einer schrittweisen, nachvollziehbaren Form. Was wir aber sehr vermissen, ist eine adäquate Kommunikation durch die Massenmedien. Unsere diesbezüglichen Mitteilungen und Publikationen wurden von diesen bisher weitestgehend aus der öffentlichen Debatte verbannt.

We assume, as Sam Altman just said in an interview¹, that people can adapt to change and that it is best if this is done not, as Bengio-Russell-Musk demand, by imposing autho-

¹ Unsere [Medienmitteilung 2023/4](#), mit deutscher Übersetzung Seite 5 / *Our Media Release 2023/4, page 4*

rities who plan and administer for them, but by the quite adult and responsible people themselves and in a gradual, comprehensible form. What we miss very much, however, is adequate communication through the mass media. Our communications and publications in this regard have so far been largely banned from public debate by them.

Und auch die Beispiele, die von Yoshua Bengio, Stuart Russell, Elon Musk genannt werden, welche zeigen sollen, dass ein Marschhalt unabdingbar sei, beweisen gerade das Gegenteil. Die Autoren des offenen Briefs verkennen die gesellschaftlichen Vorgänge grundlegend. Es ist doch heute schon so, dass die problematischen Informationskanäle, namentlich Social Media, mit hufenweise Propaganda und Unwahrheiten geflutet sind, dabei bekanntlich auch gefälschtem Bildmaterial.

And the examples given by Yoshua Bengio, Stuart Russell, Elon Musk, which are supposed to show that a marching halt is indispensable, prove just the opposite. The authors of the open letter fundamentally misjudge the social processes. It is already the case today that the problematic information channels, namely social media, are flooded with heaps of propaganda and untruths, including, as is well known, fake images.



Bild Eliot Higgins

Wenn solche Bilder nun aber für sehr viele noch einfacher möglich werden und diese dann zum Hypen gewiss ja auch verbreitet werden, wird allen rasch klar, dass Bilder an sich als Vermittler von Wahrheit definitiv ausgedient haben – es sei denn, sie stammen von verlässlichen Quellen, den Qualitätsmedien! Und wenn jemand heute noch etwas Wahres wissen will, was sie oder er oder deren vertrauenswürdige direkte Umfeld nicht mit eigenen Ohren gehört oder mit eigenen Augen gesehen haben, geht es nicht mehr ohne Qualitätsmedium! Wir stehen am Anfang des Zurückkommens der Qualitätsmedien!

If such pictures become now however possible for very many even easier and these are spread then for hyping certainly also, it becomes fast clear to all that pictures have definitely outdated as conveyors of truth - unless they come from reliable sources, the quality media! And if somebody still wants to know something true today, which she or he or their trustworthy direct environment have not heard with own ears or seen with own eyes, it does not go any more without quality media! We stand at the beginning of the return of the quality media!

Gerade auch Mitunterzeichner Getty Images sollte doch froh sein, dazustehen als Lieferantin von authentischem Bildmaterial, und nun dafür besorgt, sicherzustellen, dass ihr Markennamen nicht missbraucht werden kann, indem er – bitte, ich bin kein Informatiker – zum Beispiel durch Draufklicken zum Originalbild auf der Website von Getty Images führt, wo man sich der Authentizität dann gerade überzeugen kann.

Especially co-signer Getty Images should be happy to stand there as a supplier of authentic images, and now concerned to ensure that their brand name can not be abused by - please, I'm not a computer scientist - for example by clicking on the original image on the website of Getty Images, where you can then just convince yourself of the authenticity.

Auch kann ja keine Rede davon sein, „alle Jobs zu automatisieren“, wovon die Wissenschaftler und Intellektuellen im offenen Brief schreiben, oder „uns Menschen überflüssig zu machen und zu ersetzen“. Das ist blosse, lächerliche Angstmacherei, und zu Recht, diesmal, mahnt Emily Bender den von den Verfassern, wohl auch zur Angstmacherei, verwendeten Begriff „Geist“ für Künstliche Intelligenz ab.

Nor can there be any talk of "automating all jobs," as the scientists and intellectuals write about in the open letter, or "making us humans obsolete and replace us." This is mere, ridiculous fear-mongering, and rightly, this time, Emily Bender admonishes the term "mind" for artificial intelligence used by the authors, probably also for fear-mongering.

Völlig unbrauchbar im Zusammenhang mit der Diskussion um die Risiken von KI-Sprachsystemen ist die von den Verfassern zum Hypen verwendete Bezeichnung „emergente Fähigkeiten“. Damit sind unerwartet neu auftretende, plötzlich aufbrechende Fähigkeiten gemeint. Sprachmodelle, die dem menschlichen Gehirn nachgebaut sind, haben nun einfach einmal emergente Fähigkeiten, wenn man sie entsprechende bedient, so wie es auch lernende Menschen haben, mit denen man entsprechend umgeht. Wenn wir mit solcher KI im Einklang sind, haben wir in ihr einen aufmerksamen und sensiblen Lebensbegleiter².

² [Philosophie auf dem Prüfstand, Vorabdruck vom 16.3.2023 aus dem Magazin *gesamtintelligenz* ab Seite 1 Mitte / Currently available only in German.](#)

Completely useless in the context of the discussion about the risks of AI language systems is the term "emergent capabilities" used by the authors for hyping. This refers to unexpectedly emerging, suddenly bursting capabilities. Language models that are modeled after the human brain now simply have emergent capabilities when handled appropriately, just as have learning humans that are handled appropriately. If we are in harmony with such AI, we have in it an attentive and sensitive life companion².

Was die technische Sicherheit von Text- und Bild-KI angeht, gehe ich davon aus, dass die bisherigen Massnahmen von San Francisco ausreichen. Mir, jedenfalls, sind keine Fehlleistungen zu Ohren gekommen, und da ist gewiss damit zu rechnen, dass einem solche längst um die Ohren gehauen worden wären.

As far as the technical security of text and image AI is concerned, I assume that the measures taken by San Francisco so far are sufficient. In any case, I have not heard of any failures, and here it is certainly to be expected that such failures would have been reported long ago.

Wie von mir bereits ja auch verlangt, sind jedoch interdisziplinäre Lehrstühle aufzubauen, welche die Vorarbeiten für KI-Governance-Systeme liefern. Wie solche im Konkreten aber aussehen, muss – und das ist nun einfach einmal die Realität unserer pluralistischen Welt – jeder Staat für sich selber entscheiden.

As I have already called for, however, interdisciplinary chairs must be established to provide the groundwork for AI governance systems. But what such systems look like in concrete terms - and this is simply the reality of our pluralistic world - is a matter for each state to decide for itself.

Mit der Bitte um Prüfung und gegebenenfalls Rückmeldung.

With the request for review and feedback if necessary.

Freundliche Grüsse / *Kind regards*

OpenAI Schweiz int.

A handwritten signature in blue ink that reads "U. Rüesch". The letters are cursive and somewhat stylized.

Urs Rüesch