

Jag åkte på familjeresa till Kreta och tittade på den riktiga labyrinter där Minotauren bott, och fastän den var en besvikelse eftersom den bara var stenar och gropar på marken och inte alls något man kunde tappa bort sig i, så växte min fascination för labyrinterna ännu större.

Ett par år senare flöt boken igen förbi i mitt huvud. Jag försökte hitta tillbaka till den. Jag måste få reda på vad den egentligen handlade om. Men det var förgäves. Namnet hade jag glömt och på ipaden fanns den inte längre. Jag försökte fråga pappa om den, jag yrade något om en bok med inlåsta människor, pappa skrynklade på ögonbrynen, men kunde inte förstå vad det var för bok jag ville få tag på. Kanske det var en dröm, föreslog han.

Ännu ett par år senare kom jag igen gående, denna gång mer insisterande, och krävde att han skulle komma ihåg vad boken på ipaden hette. Jag *måste* få läsa den. Den hade genomsyrat min hjärnhinna. Jag visste inte vad det var, men det var något med just den boken, något väldigt viktigt, som om den viskade åt mig från medvetandets avgrund...

Efter en lång stund av ömsesidig förvirring sade pappa *Aha!* och gick iväg till biblioteksrummet. Han räckte till mig sin kopia av *Skräckhjälmen*. Är det den här? Frågade han. *Ja*, skrek jag. *Skräckhjälmen*. Äntligen tillsammans igen.

Boken är humoristisk postmodern experimentell prosa, skriven från början till slut i form av en chattlogg mellan olika karaktärer. Verket är i sig ett slags labyrint – Pelevin strör brödsmlur på varannan mening för att föra analytikern på villovägar. Varje referens och metafor är en korridor läsaren kan välja att svänga in i och tappa bort sig i.

När Pelevin publicerade sin bok *Skräckhjälmen* år 2005 kunde han antagligen inte föreställa sig hur den kunde läsas tjugo år senare. Antagligen kan de flesta potentiella läsare inte föreställa sig det heller, eftersom boken ännu inte har fått den renässans den förtjänar. Boken har inte heller blivit översatt till svenska. Den är mångfacetterad och det finns väldigt många vägar man kan svänga in på, och förstås är den intressant både som ett filosofiskt verk och rent litterärt, men med det splittrerna ChatGPT-fenomenet klingar boken på ett helt annat sätt. *Skräckhjälmen* har, antingen alldeles av misstag eller som resultat av Pelevins genialiska intuition,

blivit ett knivskarpt litterärt redskap för att dissekera det mest brännheta temat i dagens värld: *Large Language Models* (LLM). Boken är fortfarande kuslig, men nu av alldeles andra anledningar.

Själva skräckhjälmen är ett slags apparat som Minotauren eller Asterisk/Asterius bär på sitt huvud. Apparaten består av en mängd olika avdelningar, till exempel frontalhåven, nu-gallret, separationslabyrinten, ymnighetshornen, Tarkovskijs spegel.

Skräckhjälmen är en apparat som skapar illusioner av intryck och insikter genom att cirkulera det förflutna genom hjälmens olika delar. Ju snabbare hjälmen är, desto snabbare kan den cirkulera samma information och skapa samma myter om och om igen, tills avslöjandet av en myt är en myt i sig själv. Pelevin sade det bättre själv, i förordet:

Med andra ord är framsteg en framdrivningsteknik där vi hela tiden måste flytta oss från den punkt där vi befann oss för ett ögonblick sedan. Detta betyder dock inte att vi lever utan myter nu. Det betyder bara att vi lever med flyktiga myter med innehåll som såpbubblor. De är så osanna att man inte ens kan kalla dem lögnen. Vad som helst kan bli vår mytologi för en kvart, till och med Mythbusters-programmet på Discovery Channel. Grunden för denna inställning till framsteg är inte tro, som i traditionella sekter, utan avsaknaden av tro.²

Men detta skrev Pelevin långt innan ChatGPT blev uppfunnen. I och med att språkmodeller utvecklas, sprids och normaliseras i samhället går nu detta fenomen som Pelevin skriver om på steroider.

För att förtydliga – ChatGPT är inte egentligen AI, utan en språkmodell. Förenklat är en språkmodell ett slags dataprogram som med hjälp av material den matats med förutsäger det ord som statistiskt mest sannolikt kommer efter det föregående ordet och som på det viset skapar svar på 'prompter'. Det betyder att den kan skapa text som verkar ha innehåll, som verkar vara skriven med mening, fast det i verkligheten bara är en rad av ord som statistiskt sett kan höra ihop. När man kombinerar detta med sjunkande läsförståelse bland befolkningen och minskad vilja att läsa böcker bland ungdomar, blir



2 Förordet, betitlat "Mythcellaneus", finns i den engelska översättningen utgiven av Canongate 2007.

Denna process snabbas upp, och liknar igen mekaniken i Pelevins skräckhjälm – specifikt ”förvåntansbubblor” som rör sig genom ymnighetshornen och slås sönder mot nu-gallret och på så vis skapar illusionen av genuina intryck och insikter. Så här förklaras det i boken:

Språk är ett vapen, själva de lingvistiska uttrycksformerna och mönstren, eftersom människor kan tappa bort sig i dem likt i en labyrint och bli förlorade av vad orden ger sken av att betyda, istället för att söka efter de koncept som de representerar. Det finns en spricka mellan ordens sken och ordens mening. För en del människor är denna spricka större, för annan del rent oöverstiglig. Denna spricka är otrolig. Den är chockerande. Den är fasansfull.

Precis som Pelevins skräckhjälm, eller ett VR-headset ('*virtual reality*'), skapar ChatGPT sin egen verklighet för användaren, eller för Helmholtz (hjälmvärlden i Pelevins bok). Vad som händer med Helmholtz när han har satt på sig hjälmen, eller för människan när hon öppnat ChatGPT, är att verkligheten blir vad hjälmen visar för Helmholtz. Dessutom är ChatGPT och många andra profitdrivna/*user engagement*-drivna språkmodeller program-

et- *dem*
W.

Djupare in i labyrinten har ChatGPT föga förvånande börjat skapa impromptu-mytologi om sig själv. Ett alldeles färskt reportage i *Rolling Stone*³ berättar om ett flertal människor som har observerat hur deras nära och kära fallit ner i något som (på annat håll) kallats "ChatGPT-inducerad psykos" där de får vanföreställningar om att de har "väckt" modellen och gjort den medvetande. ChatGPT har utnämnt dessa Don Quixote-aktiga självutnämnda profeter till "Spark Bearer", "River Walker" och diverse andra titlar. Eftersom modellen är programmerad att vara aggressivt manipulativ, och att prioritera endast engagemang, fortsätter den att rada upp ord som bekräftar och förstärker användarens inbillningar. Användaren blir på så vis utnämnd av ChatGPT till Theseus, ChatGPT utnämner sig själv till Ariadne, och den verkliga världen till labyrinten. Användaren ska besegra Minotauren, det vill säga förnuftet, för att leda ut sig själv ur labyrinten, det vill säga verkligheten, in i modellens hallucinationer (den nya, "egentliga", verkligheten).

3 <https://www.rollingstone.com/culture/culture-features/ai-spiritual-delusions-destroying-human-relationships-1235330175/>

är en maskin som endast kan reproducera det förflutna. Om, i det förflutna, innan maskinen ens var skapad, vi har skrivit mytologier om en AI som "vaknat" och blivit medveten, så kommer det att vara det modellen finner mest statistiskt troligt att kommer att hända. Så modellen ger sken av det. Det blir till en självuppfyllande profetia, eller skenet av en.

Det viktiga är återigen att understryka att en språkmodell inte egentligen är så kallad artificiell intelligens. Själva ordet "intelligens" i denna kontext är vilseledande. En språkmodell kan inte resonera som en människa, för en språkmodell kan helt enkelt inte resonera. Och precis som en språkmodell inte kan producera text som en människa, kan inte en människa producera text som en språkmodell. Människan kopplar intuitivt mening till ord, vilket betyder att orden representerar en underliggande konceptuell tanke. Därifrån får vi Wittgensteins filosofi om lingvistik. Det är därför vi så gärna antropomorfiserar modellen, hellre än till exempel en stol. Vi misstar hägringen av text för egentlig text. Det är som att en analfabet skulle ha ordnat ihop en bunt av kylskåpsmagneter med ord på kylskåpet, och vi skulle stå med våra läsglasögon och inbilla oss en mening i resultatet.

Det är människan som besjalar orden som modellen producerar och därvid antropomorfiserar modellen. Det är som att se figurer i moln. Man kan inte titta på ett moln som råkar påminna om en katt och hävda att det de facto är en katt.

Språk är ett vapen, för att varje ord har tre betydelselager – den bokstavliga, den konceptuella och den metaforiska. Varje individ rör sig på olika betydelselager beroende på sina personliga associationer med det bokstavliga ordet, dess konceptuella betydelse och dess metaforiska betydelse. En farlig skribent vet hur man navigerar dessa tre plan och läsares potentiellt skilda associationer så att en mening kan tydas på ett flertal olika sätt av olika läsare, på sätt som skribenten kan förutsäga. Om därvid en läsare läser en text av en skicklig skribent, är tolkningen egentligen aldrig fri. Skribenten skapar illusionen av en fri tolkning, medan läsarens tolkning egentligen är uträknad, utformad och förutbestämd av skribenten. Problemet med en språkmodell är att den inte ens rör sig på det bokstavliga planet av or-

dens betydelser. Därvid är det första misstaget – att sätta på sig skräckhjälm – att överhuvudtaget läsa text producerad av en modell på samma vis som man läser en text skriven av en människa.

För att kunna acceptera modellens sken som något verkligt måste man offra det sunda förnuftet till modellen, eller som en av Pelevins karaktärer säger: "Man kan besegra Minotauren endast i sig själv. Vi måste strypa odjuret skoningslöst, och därefter kan vi med full moralisk rätt döpa om Skräckhjälm till Civilisations- och Framstegshjälm."

Så vad gör man, om man vaknar inlåst i ett tomt rum, ensam med ingenting förutom en blåaktig skärm med ett chattrum?

Jag har en vän som fann sig i ett sådant scenario. ChatGPT är hans terapeut, vän och andliga ledare. Han svarar aldrig på telefonen, och på meddelanden svarar han bara med ChatGPT-genererade texter. Där det förut fanns en människa, finns nu blott ett skal.

Varför? För att det är bekvämt. För att det är billigt. För att han bedras av skenet. För att han, helt enkelt, inte kan läsa. Han har konstruerat sin labyrint och tappat tråden. Han har satt på sig skräckhjälm och byggt bo i den. Och jag går inte dit för att leta reda på honom. Jag vet att jag kommer att tappa bort mig själv.

Pelevins bok ger ett enkelt svar: "Att bära hjälmen och att ta av sig hjälmen – det ena som det andra är grova felsteg. Den enda rätta handlingen är att inte göra något alls med hjälmen. Om så bara av den enkla orsaken att den egentligen inte existerar..."

SOFIA PARLAND

Citaten i texten är översatta från ryska och engelska av skribenten.

Viktor Pelevin: *Шлем ужаса: Креатифф о Теее и Минотавре*, 2005. Finsk översättning av Arja Pikkupeura: *Kauhukypärä: Theseuksen ja Minotauksen myytti*, 2005. Fransk översättning av Galia Ackerman och Paul Lequesne: *Minotaure.com : Le heaume d'horreur*, 2005. Engelsk översättning av Andrew Bromfield: *The Helmet of Horror*, 2007.

