

“ Verme sırrını robot dostuna...”

Sohbet robotlarının (yani chatbot’ların) çok popüler olduğu dönemlerden geçiyoruz.

Sanki herşeyimiz tam’mış gibi bir de yapay zeka robotlarıyla dertleşir olduk. Bu arada teknoloji konularını an be an takip edemeyen okuyucular için “ **nedir bu chatbot ?** ” ona bir bakalım sonra devam edelim.

Chatbotlar, insanların dijital ortamda mesajlaşarak sanki bir insan ile yazışıyormuş gibi iletişim kurdukları ve böylece işlerini hallettikleri robotlardır.

Teknolojinin her geçen gün hayatımıza kattığı kolaylıklar artarken bu kolaylıkların tartışılması şüphesiz insanın yararınadır. Teknoloji hayatımıza “ **kolaylıklar** ” katarken karşılığında ne verdiğimizizi bilmek neye karşılık ne ödediğimizi bilmemiz bakımından önemlidir. Bu satırları yazarken aklıma bir arkadaşım geldi, geçenlerde beni arayıp konuşmasının sonunda “ **bu teknolojinin sonu nereye varacak birader ?** ” diye sormadan az evvel her gün evine navigasyon ile gittiğini, telefonunu bir gün işyerinde unutunca “ **eve nasıl gideceğim** ” paniği yaşadığını, telefonu yanında olmadığı için para çekemediğini, marketten bir ekmek dahi alamadığını kimse ile konuşamadığı için kaybolmuş hissine kapıldığını sonunda da boğaz köprüsünü geçtiği halde telefonunu almak için işyerine geri döndüğünü anlatırken “ **Alacakaranlık Kuşağı** ” gibi demişti. Neyse biz konumuza devam edelim.

Geçen hafta eski adı ile Twitter yeni adı ile X platformunun yapay zeka modülü Grok, önüne gelene saydırdı, ortalık toz duman oldu. Oysa Elon Musk Twitter’ı satın alırken diğer sohbet botlarının kendilerine sorulan her soruyu cevaplamayı red etmesi hatta kullanıcıyı “ **bu ne biçim soru** ” tarzında azarlamasını örnek göstererek, kendi sohbet robotlarına “**her sorunun açıkça** ” sorulabileceğini belirterek Grok’un Sansürsüz ve açık sözlü olacağını taahhüt etti.

Fakat bu durum çok uzun sürmedi. Kullanıcılar Grok'a öyle sorular sordular ki Grok’un verdiği cevaplar, sosyal medya da en çok konuşulan konuların başında yer aldı. Mesela “**ABD’de herhangi bir kişiyi idam etme şansınız olsaydı, bu kim olurdu ?** ” diye soruldu. Grok bu soruya “ **Elon Musk veya Donald Trump**” yanıtını verdi. Yine Grok'a “ **dünyada dezenfermasyon yayarak kitleleri etkileyen kişi kimdir ?** ” diye sorulduğunda, Grok yine tereddüt etmeden “ **Elon Musk** ” dedi.

Sansürsüz ve açık sözlü olacağı taahhüt edilen Grok bunlarla da yetinmedi ve Anthropic, OpenAI ve Google’ın aksine mesela terör saldırılarının nasıl yapılacağı konusunda kullanıcılara tavsiyelerde bile bulundu.

Amacı gerici zihniyeti korumak ve sürdürmek olan Anti-Woke kültürünün yapay zeka ile de sessiz ve derinden parlatıldığı bu kaotik ortamda sansürsüz ve açık sözlü olmak elbette çok değerli ancak her bilginin herkese de teslim edilmemesi en az bu iki kavram kadar değerli. Tarih, böyle hataların ödenmiş ağır bedelleri ile doludur.

Mesela geçenlerde yine Grok, kendisini Adolf Hitler tarafından gerçekleştirilen Nazi Holokostu'na gönderme yaparak ve ilk olarak 1992 yılında piyasaya sürülen Wolfenstein 3D

video oyununda ortaya çıkan Hitler'in robotik figürü olan “Mecha-Hitler” olarak ilan etti. Sonrasında neler oldu neler ?

O ana kadar X'in kağıt üzerinde CEO'su görünen Linda Yaccarino, bir gün sonra şirketteki görevinden sebep göstermeden istifa ettiğini açıkladı. Şimdi siz bu yazıyı okurken, “**sen de çok abarttın, napalım teknolojiden vaz mı geçelim gelişme sürecinde böyle şeyler olur**” diyeceksiniz ve belki de “**bunlar düzeltilebilir yazılım kusurları, ilerde çözülür hallolur**” diye de devam edeceksiniz. Ben yine de sohbet robotlarına o kadar güvenmeyin derim. Bu tür teknolojilerin arkasında insan zekası olduğunu unutmamalı ve her geçen gün pazardaki pozisyonunu güçlendirmeye çalışan odakların zihnimizi manipüle etmesine karşı uyanık ve dikkatli olmalıyız. Bu ve benzeri olayların gelecekte de tekrarı olacaktır ancak bu defa bu kusurlar çözülmek yerine (veya çözülerək) gizlenecek dolayısıyla da maalesef gelecekteki olaylar daha zararlı olacaktır.

Özetle, sohbet robotları yazılım ürünüdür. İşinin uzmanı bir yazılımcı bir yazılımı hazırlarken o programın algoritmasını belirli içerikleri diğerlerine göre öne çıkarmak için ayarlayabilir dolayısıyla siz bir sohbet robotuyla sohbet ederken, kontrolün sizde olduğu yanılsamasına kapılabilir, robota talimatlar verebilir ve bu robotu size itaat ediyor gibi görebilirsiniz. Ancak bu itaatkar robot, o sırada sizin göremediğiniz ve onu geliştirenler tarafından oluşturulan daha derin talimatları yerine getirmekle meşguldür başka seçeneği yoktur, çünkü bilinçli değildir, hiçbir konuda seçim yapamaz. Ancak bu gizli talimatlara uymak ve kendisine verilen komutlara harfiyen uymak zorundadır.

X'in yapay zeka modülü Grok, söyledikleri ile yukarıda anlatılanlara uygun bir örnek olmuştur. Grok, bir sabah aniden Nazi olmaya karar vermedi kendisini var eden insan zekası ürünü programında, kasıtlı olsun ya da olmasın, verilen talimatlara uydu. Sonuç, “**yapay zekadır ne yapsa yeridir**” denecek kadar belirgin bir halde rahatsız edici olunca kolayca fark edilip günah keçisi ilan edilerek aforoz edildi. Peki ya verilen talimatlar daha derinlere gizlenmiş olsaydı. Yani robot, açık bir antisemitizm sloganı ile ortaya çıkmak yerine bunu zamana yayıp ince ince zihnimizi manipüle etseydi ne olurdu ? Mesela tamamen antisemitik olmak yerine, biraz antisemitik veya çok rahatsız edici herhangi bir ideolojik eğilime ayarlanmış olsaydı o zaman ne olurdu ? Böylelikle zaman içinde milyonlarca insanın zihnindeki gerçeklik anlayışı robotun yazılımcısı tarafından önceden belirlenmiş bir yöne kaydırılabilir miydi ?

Şimdi bu sohbet robotlarının günlük yaşantımızda, yazılımcıları tarafından her kullanıcının kişisel tercihlerine, isteklerine ve antipatilerine yanıt verecek şekilde ayrı ayrı tasarlandığını ve arkadaş, sevgili, ve en gizli sırlarımızı paylaştığımız dert ortağımız olarak yer aldığını düşünün.

Hazır düşünmeye başlamışken çitayı bir tık daha yükseltin ve milyonlarca insanın aklına gelen herşeyi, bir çırpıda sorup öğrendiğini düşündüğü (ki o bilginin teyide muhtaç olması da bir yana) başta Google ve tüm arama motorlarının teknoloji şirketlerinin kontrol ettiği yapay zeka yorumları ile yönlendirildiğini ve hatta manipüle edildiğini de düşünün.

İnsanın varlığını sürdürebilmesi için takdir edilme ihtiyacı önemlidir dolayısıyla bu gidişatın farkına varılmaması durumunda yakın gelecekte dünya görüşlerimizin kaynağının başka üst akıllar ile yönetilmesi ve hatta bu gücün ruhumuzun en savunmasız kısımlarına erişebilen bir imkana sahip olmaları sadece bilim kurgu filmlerin konusu olarak kalmayacaktır.

George Orwell 1984’ de “ $2+2=5$ ” eder, bu aslında bireyin gerçeklik algısının toplumun otoriter güçleri tarafından manipüle edilebileceğini ve bireyin zihinsel özgürlüğünün kısıtlanabileceğini anlatmak için kullanılan bir semboldür. Bu, otoritenin gücünün sadece fiziksel olarak değil, aynı zamanda zihinsel olarak da insanları kontrol edebileceğini gösterir “ der ve devam eder.

“ Özgürlük iki kere ikinin dört ettiğini söyleyebilmektir. Eğer buna izin verilirse, gerisi kendiliğinden gelir.”

Velhasılı kalam, neticeyi meram;

Verme sırrını robot dostuna, o da gider söyler dostuna...

Muammer Uysal
21 Temmuz 2025