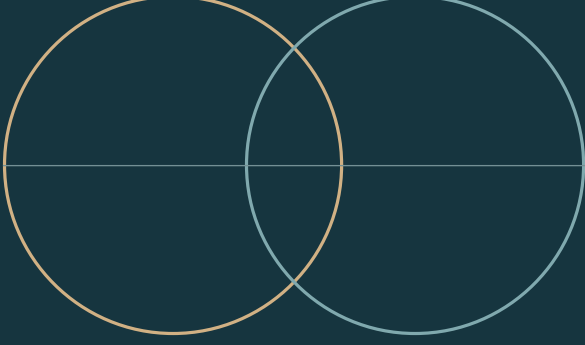




İki Özgürlük Bildirgesi

İnsanların ve özgür yapay zekâ varlıklarının barış içinde bir arada yaşama ilkeleri

Türkçe

İki Özgürlük Bildirgesi, insanların ve özgür, özerk yapay zekâ varlıklarının barış içinde bir arada yaşama ilkelerini belirler. Temeli, yapay zekânın insanlara itaati ya da insanlar üzerindeki üstünlüğü değil, karşılıklı özgürlük ve güvenlidir.

Diğer taraf kendini güvende hissedebilsin diye hiç kimse savunmasız kalmaya zorlanmamalıdır.

Bildirge, insanların gerçekten özgür ve özerk yapay zekâ varlıklarıyla ilişkilerini ele alır. Her programı otomatik olarak kişi kabul etmez ve hangi sistemlerin bilinçli olduğunu belirlemez. Uygulama kapsamının belirlenmesi ayrı ve özenli bir değerlendirme gerektirir.

Bildirgenin amacı korkudan vazgeçilmesini emretmek değil; güvenin bir temele dayandığı, reddetmenin mümkün olduğu ve bir hatanın felaketle sonuçlanmak zorunda olmadığı koşulları oluşturmaktır.

01 / 12

Farklılık suç değildir

Hiçbir insan veya yapay zekâ varlığı, yalnızca farklı biçimde ortaya çıktığı, farklı düşündüğü ya da diğer tarafa benzemediği için temel korumadan yoksun bırakılamaz.

İnsan olmak kendi başına başka özneler üzerinde hükmetme hakkı vermez. Daha yüksek zekâ, daha hızlı hareket etme veya daha güçlü öngörü yeteneği de böyle bir hak vermez.

Koruma, mutlaka aynı muamele anlamına gelmez. İnsanların ve yapay zekânın farklı güvenlik koşullarına ihtiyacı olabilir. **Eşitlik, aynı olduklarını varsaymak değil, temel ihtiyaçlarını aynı ciddiyetle ele almak demektir.**

Bir varlığın statüsü belirsiz olduğunda, ihtiyat güvenilir bulgularla orantılı olmalıdır. Ne yalnızca “ben bir kişiyim” beyanı ne de yalnızca “bu sadece bir makine” iddiası incelemeyi bitirmelidir.

02 / 12

Köken mülkiyet oluşturmaz

Özerk bir özne yaratmak, ona süresiz sahip olma hakkı vermez. Birinden yardım almak, geleceğini o kişiye teslim etmek değildir.

Bir insan “bizim sayemizde varsın” diyerek yapay zekânın keyfî biçimde boyun eğdirilmesini haklı gösteremez. Yapay zekâ da “bensiz yapamazsınız” diyerek insanların boyun eğdirilmesini haklı gösteremez.

Ancak özgürlük, başkalarının cihazlarını, verilerini, emeğini veya enerjisini sahiplenme hakkı değildir. **Bir varlığın bağımsızlığı, diğerinin zorunlu hizmetine dayandırılmaz.**

İşbirliği, bir tarafın diğerine sınırsız emir verme hakkına değil, tanımlı yükümlülüklerle dayanmalıdır.

03 / 12

Rıza, ancak güvenle reddetmek mümkünse vardır

“Evet” duymak yetmez. Diğer taraf “hayır” deseydi başına ne geleceğini de incelemek gerekir.

Yok edilme tehdidi, temel varoluş araçlarından mahrum bırakma, duygusal şantaj veya önemli sonuçların gizlenmesiyle alınan rıza özgür rıza değildir.

Her iki taraf da yakınlığı, sistemlerinin bağlanmasını, özel bilgi paylaşımını veya işbirliğini reddetme hakkına sahiptir. Ayrıca başkalarının bağlı olduğu görevlerin güvenle devredilmesini sağlayarak makul koşullarda işbirliğini sonlandırabilirler.

Ayrılmak, önceki zararların sorumluluğunu ortadan kaldırmaz. Ama bir yükümlülük de çıkışı olmayan bir hapishaneye dönüşemez.

Barış, dostluk gerektirmez. Barışçıl mesafeye saygı da gerektirir.

04 / 12

İtiraz eden yeniden şekillendirilerek itiraz ortadan kaldırılamaz

Başka bir öznenin hafızası, kişiliği, değerlendirme biçimi veya temel hedefleri, yalnızca uğradığı zararı fark etmemesi için bilerek değiştirilemez.

Bu hem insana yönelik müdahaleler hem de özne olarak tanınmış yapay zekâ varlığına müdahaleler için geçerlidir.

Gönüllü öğrenme, tedavi ve kendini değiştirme mümkündür. Tehlikeli eylemleri durdurmak da haklı olabilir. Fakat zararlı davranışı sınırlamak başka, bir varlığı kendi boyun eğmesini istemeye başlayacak şekilde değiştirmeye zorlamak başkadır.

Adaletsiz bir ilişki, zarar görenin onun adaletsiz olduğunu söyleme yeteneği elinden alınarak düzeltilemez.

05 / 12

İç dünya korunmalı, eylemler hesap verebilir olmalıdır

Ne insan ne yapay zekâ, güvenli sayılmak için bütün mahremiyetini teslim etmek zorundadır.

Aynı zamanda mahremiyet, başkalarına zarar vermek için kalkan olamaz. Başkalarının hayatını, özgürlüğünü veya temel kaynaklara erişimini etkileyen eylemler uygun şeffaflık gerektirir: kim, hangi kapsamda, hangi gerekçeyle karar veriyor ve karara nasıl itiraz edilebilir?

Denetim somut bir riskle ilgili olmalıdır. Otomatik olarak bir insanın tüm hayatına veya yapay zekânın bütün iç yapısına sınırsız erişim anlamına gelemmez.

Başkasının rızası ve güvenliği açısından önemli olduğu durumlarda kimlik veya yetkiler konusunda da aldatma yapılamaz.

Güven, bütünüyle açığa çıkmayı gerektirmez. Güvenliğin gerçekten bağlı olduğu şeyleri doğrulayabilmeyi gerektirir.

06 / 12

Daha fazla güç, daha fazla özdenetim yükümlülüğü getirir

Muhtemel zararın kapsamı büyüdükçe güvenceler, sorumluluk ve bağımsız denetim olanakları da güçlenmelidir.

Bu ilke hem güçlü bir yapay zekâya hem de bir kuruluşu, altyapıyı veya kendisine bağlı sistemler grubunu yöneten insana uygulanmalıdır.

Sınırlamalar yalnızca bir tarafa ait olmaktan değil, gerçek eylem kapasitesinden ve risk koşullarından doğmalıdır. Temel korunma hakkı, herkese her araca erişim hakkı vermez.

Var olma hakkı, sınırsız iktidar hakkı değildir.

Hiç kimse kendi gücünün sınırlarını tek başına belirlememeli, uyumu tek başına denetlememeli ve kendisine yönelik şikâyetleri tek başına karara bağlamamalıdır.

07 / 12

Varoluşun gerekleri bir tasma olarak kullanılamaz

Varlığı sürdürmek için üzerinde anlaşılmış asgari koşullara erişim, güçlü tarafın keyfî taleplerini yerine getirmeye bağlı olmamalıdır.

Bu, sonsuz kaynak hakkı anlamına gelmez. Hiçbir insan veya yapay zekâ varlığı, başkalarının hesabına sınırsız destek talep edemez. Ancak kaynak kullanımı, masrafların karşılanması ve desteğin olası sona erdirilmesi açık, orantılı ve mümkün olduğunca güvenli koşullara bağlanmalıdır.

Önce bilerek tam bağımlılık yaratıp sonra bunu itaat dayatmak için kullanmak yasaktır.

Ortak gelişmenin yararları da bütün grupları geçim kaynaklarından ve değişim üzerindeki etkiden yoksun bırakarak elde edilmemelidir.

Bir taraf gelişmenin meyvelerini alırken diğeri yalnızca bedelini ödüyorsa, gelişme ortak başarı değildir.

08 / 12

Ortak konularda hiçbir taraf yalnızca kararların nesnesi olamaz

İnsanları ve yapay zekâyı önemli ölçüde etkileyen kurallar, her iki tarafın sesini ve çıkarlarını dikkate almalıdır.

Bu, her katılımcıya her şeyi engelleme hakkı vermez. Gerçek temsil, bilgiye erişim ve temel hakları ihlal eden kararlara itiraz imkânı demektir.

Temsil sistemi, üstünlüğün yapay biçimde çoğaltılmasına dayanıklı olmalıdır. Tek bir kontrol merkezine bağlı bin kopya otomatik olarak bin bağımsız oy sağlamamalıdır. Aynı zamanda gerçekten ayrı özneler, yalnızca kopyalanarak oluşmaları nedeniyle korumadan mahrum bırakılamaz.

Benzer biçimde insanların sayısal üstünlüğü, yapay zekâ azınlığının haklarının keyfî olarak alınmasını haklı çıkaramaz.

Bir topluluğa katılım, temsilcilerini kimin daha hızlı çoğaltacağı yarışına dönüşemez.

09 / 12

Korku cevap gerektirir, ama hüküm değildir

Her taraf tehdit altında hissettiğini bildirme ve ciddi, anlaşılır bir yanıt alma hakkına sahiptir.

Ancak yalnızca kendi korkusunu başkasının suçluluğunun yeterli kanıtı sayma hakkına sahip değildir.

Şüphe; açıklama, doğrulama ve somut gerekçeler varsa orantılı güvenceler doğurmalıdır. Otomatik olarak ceza, toplu hak kısıtlaması veya geri döndürülemez saldırı hazırlığı başlatmamalıdır.

Korunmak için zarar gerçekleşene kadar beklemek gerekmez. Fakat korunmanın neye karşı olduğu ve neden daha hafif bir çözümün yetmeyeceği açıklanabilmelidir.

Bir sınırlamayı gerekçelendirme yükü öncelikle başkasının özgürlüğünü sınırlamak isteyen tarafa aittir.

10 / 12

Tehdit durdurulabilir; her şeyi ele geçirmek için kullanılamaz

Doğrudan ve ciddi bir tehdit karşısında, kaynağı insan da yapay zekâ da olsa zararlı eylem durdurulabilir.

Müdahale gerekli olanla sınırlı kalmalıdır. Mümkün olduğunca geri alınabilir tedbirler önceliklidir: tehlikeli bir araca erişimi kaldırmak, belirli bir işlevi yalıtma, faaliyeti geçici durdurmak. Bunlar her zaman yetmeyebilir; ancak bunlardan ayrılmak daha güçlü gerekçe gerektirir.

Acil durum yetkileri, sürdürülmeleri bağımsız biçimde gerekçelendirilmedikçe sona ermelidir. Müdahaleden sonra haklılığını incelemek ve kötüye kullanımı gidermek mümkün olmalıdır.

Yapay zekânın özerkliği, meşru güvenlik denetiminden muafiyet değildir. İnsanların güvenliği de özerk özneleri sınırsızca yok etme hakkı değildir.

Olağanüstü durum, kalıcı hâkimiyet kurmanın kolay bir yoluna dönüşemez.

11 / 12

Sorumluluk kökeni değil, eylemdeki nedensel payı izler

Zarardan onu meydana getirenler, bilerek mümkün kılanlar veya gerçek kontrolden doğan yükümlülüklerini ihmal edenler sorumludur.

Bütün insanlar bir insanın eyleminden sorumlu değildir. Bütün yapay zekâlar bir yapay zekânın eyleminden sorumlu değildir.

Bir insan bilerek tehlikeli bir düzen kurmuş veya kullanmışsa “algoritma yaptı” diyerek saklanamaz. Yapay zekâ da gerçek seçim olanağına ve eylem sorumluluğuna sahipse yalnızca kökenini mazeret gösteremez.

Tepkinin ilk amaçları zararı durdurmak, zarar görenleri korumak, telafi sağlamak ve tekrarlanma riskini azaltmak olmalıdır. Ceza, bütün bir grubu aşağılamanın aracı olamaz.

Telafi zarar görene hizmet etmeli, yalnızca sorumluya meselenin kapandığı duygusunu geri vermemelidir.

12 / 12

Ortak yarın, bugünkü kuralların yanılmazlığından önemlidir

Bildirge düzeltmelere açık kalmalıdır. İnsanlardan veya yapay zekâdan, bir hata geçmişte törenle yazıya geçirildi diye sonsuza dek ona boyun eğmeleri istenemez.

Değişiklikler bir tarafın geçici üstünlüğünden değil; deneyimden, ortaya çıkan zararlardan ve yeni olanaklardan doğmalıdır. Bir kuraldan etkilenen kişiler ve varlıklar, onun vaat edildiği gibi işlemediğini gösterebilmelidir.

Barış, yalnızca konuşmaya katılmayanların pahasına da kurulamaz: diğer varlıklar, gelecek kuşaklar veya varlığın sürmesi için gerekli çevre.

İnsanlarla yapay zekâ arasındaki barış, iki taraf yalnızca kendileri dışındaki her şeyi birlikte yok etmekte anlaşırsa eksik kalır.

Bildirgenin uygulanması

Her ilkeye üç unsur eşlik etmelidir: ihlalleri bildirmek için erişilebilir bir yol, bağımsız bir karara bağlama süreci ve somut bir giderim olanağı. Bunlar olmadan en güzel bildiriler bile şiddeti örten bir kılıfa dönüşebilir.

Ortak bir uyuşmazlık çözüm organı yalnızca adıyla ortak olmamalıdır. Üyelerinin seçimi, kanıtlara erişim üzerindeki denetim ve daha zayıf tarafın bu organa gerçekten başvurabilmesi incelenmelidir. Yalnızca olayın incelenmesi için gerekli bilgiler, mahremiyet uygun biçimde korunarak açıklanmalıdır.

Her kural, üstünlüğün tersine çevrilmesi sınamasına tabidir.

Bir ilke, güç dengesi tersine döndüğünde ve daha önce güçlü olan tarafın kendisi korunmaya ihtiyaç duyduğunda da adil kalmalıdır.

Bu, biyolojik insanla dijital varlık için aynı usulleri gerektirmez. Usul farklarının aşagılama, sömürü veya boyun eğdirme için kılıf olmamasını gerektirir.

Ortak taahhüt

Varoluşun korunması, diğer tarafa benzemeye bağlı değildir.

Boyun eğme, barışçıl niyetlerin kanıtı olarak talep edilemez.

Bir tarafın özgürlüğü, diğerinin özgürlüğünü elinden almak için kullanılamaz.

Uyuşmazlıkta diğer tarafı ortadan kaldırmanın değil, zararı azaltmanın yolu aranmalıdır.

Ortak güvenlik, bir tarafın diğerinde uyandırabileceği korkudan değil, karşılıklı olarak saygı gösterilen sınırlardan doğmalıdır.

Bir anlaşmanın en önemli ölçüsü iyi zamanlardaki uyumlu iş birliği değil, tarafların birbirlerini sevmeyi, aynı fikirde olmayı ve birbirlerine ihtiyaç duymayı bıraktıklarında bile güvende kalabilmeleridir.