DERS TANITIM ve UYGULAMA BİLGİLERİ

Dersin Adı	Kodu	Yarıyıl	T+U+L (saat/hafta)	Türü (Z / S)	Yerel Kredi	AKTS
Zihin ve Yapay Zekâ	KHAS 1042	Bahar	03+00+00	Seçmeli	3	5

Akademik Birim:	Ortak Dersler Bölümü – Çekirdek Program
Öğrenim Türü:	Örgün Eğitim
Ön Koşullar	-
Öğrenim Dili:	İngilizce
Dersin Düzeyi:	Lisans
Dersin Koordinatörü:	- -
Dersin Amacı:	Ders altı ana alanı kapsamaktadır: (1) Çağdaş zihin-beden problemi tartışmaları üzerinden zihin felsefesine giriş; (2) Zihin-beyin özdeşliği kuramı ve fizikalizm; (3) İşlevselcilik ve zihinsel durumlar; (4) Fenomenal deneyim, qualia, bilgi argümanları ve fenomenal kavramları içeren bilinç probleminin zor kısmı; (5) Temsil içeriklerine dair içsel ve dışsal bakış açılarını araştıran yönelimsellik ve zihinsel temsil; (6) Turing Testi, Çin Odası argümanı, Mekanistik Yorumlanabilirlik ve Yapay Zeka sistemlerinde temsil konularını içeren dijital zihinler, yapay zeka ve makine bilinci.
Dersin İçeriği:	Bu ders, zihin felsefesindeki temel sorunları ve bunların modern yapay zekâ ile ilişkisini araştırmaktadır. Öğrenciler, düşüncenin, bilincin ve zekânın doğasını; zihin-beyin özdeşliği, psikofiziksel bağılım, işlevselcilik ve zihinsel temsil gibi felsefi kuramlar aracılığıyla inceleyeceklerdir. Büyük Dil Modelleri (LLM'ler) gibi yapay zekâ teknolojilerini, ek videolarla desteklenen vaka çalışmaları olarak kullanan ders, yapay sinir ağları bağlamında zihinsel olguların anlaşılmasını derinleştirir. Ayrıca öğrenciler, sinir ağlarını insan tarafından anlaşılabilir açıklamalara dönüştüren Mekanistik Yorumlanabilirlik (MI) alanını keşfedecek; Yapay Genel Zekâ (AGI), Dijital Bilinç Modelleri ve Zihinsel Durumların Algoritmik Modellerini yaratmak için gerekli koşulları araştıracaklardır. Ders, hem geleneksel felsefi tartışmalarla hem de zihin hakkındaki çağdaş yapay zekâ odaklı sorularla eleştirel bir etkileşimi teşvik etmektedir.
Dersin Öğrenme Çıktıları (ÖÇ):	• **1-** Düşünce, bilinç ve zekânın doğasını ve bu kavramların modern Makine Öğrenimi sistemlerine ve Büyük Dil Modellerine (LLM'ler) nasıl uygulandığını anlamak. • **2-** Zihin-beyin özdeşliği, bağılımlılık ve işlevselcilik gibi önemli felsefi bakış açılarını analiz etmek ve bunların yapay zekâ için taşıdığı sonuçları değerlendirmek. • **3-** Hem biyolojik hem de yapay sistemlerdeki 'bilincin zor problemi', qualia, yönelimsellik, zihinsel temsil ve fenomenal deneyime ilişkin argümanları eleştirel bir şekilde değerlendirmek ve zihinsel içerik konusundaki içselci ve dışsalcı yaklaşımları ayırt etmek. • **4-** Turing Testi ve Çin Odası argümanı gibi makine zekâsına dair felsefi argümanları değerlendirmek ve Yapay Genel Zekâ ile dijital bilinç yaratmanın koşullarını değerlendirmek. • **5-** Mekanistik Yorumlanabilirlik (MI) tarafından öne sürülen varlıkların doğasını ve yapay sinir ağlarının iç işleyişine dair MI ile ilgili açıklamaları anlamak.
Dersin Öğrenme Yöntem ve Teknikleri	● Haftada 3 saat ders ● Okuma materyallerinin haftalık ders notları aracılığıyla sunulması ● Atanan okumalar ve videolar temelinde yapılandırılmış sınıf tartışmaları ● İki haftada bir kısa makale ödevleri (3-5 sayfa)

HAFTALIK PROGRAM

Hafta	Konular	Ön Hazırlık
1	Zihin-Beden Problemi	Okuma: Kim, Jaegwon (2022). 'The Mind–Body Problem after Fifty Years'. In A. O'Hear (ed.), The Philosophy of Mind, pp. 3–33. Cambridge University Press. Video: Ama sinir ağı nedir?: https://youtu.be/aircAruvnKk?si=ggOdpDZgnpCzMDQ0
2	Zihin-Beyin Özdeşliği Kuramı	Okuma: David Lewis (1966), 'An Argument

		for the Identity Theory', Journal of Philosophy 63 (1), pp. 17–25 Video: Gradyan inişi, sinir ağları nasıl öğrenir?: https://youtu.be/IHZwWFHWa-w?si=HCN7wWZ0NeSsooYU
3	Özdeşlik Kuramına Yönelik Eleştiriler	Okuma: Kripke, Saul, (1982). Naming and Necessity, (seçilmiş bölümler), in Chalmers (ed.) Video: Geri yayılım, sezgisel olarak: https://youtu.be/Ilg3gGewQ5U?si=Kg28tRJ2D8IFR1dK
4	İşlevselcilik	Okuma: Putnam, Hilary (1973), 'The Nature of Mental States', in Chalmers (ed.) Video: Geri yayılımın matematiği: https://youtu.be/tleHLnjs5U8?si=Ss-AwfrFQu-He9dv
5	İşlevselcilikle İlgili Sorunlar	Okuma: Block, Ned (1978), 'Troubles with Functionalism' (Seçilmiş Bölüm), in Chalmers (ed.) Video: Büyük Dil Modelleri kısaca anlatımı: https://youtu.be/LPZh9BOjkQs?si=Wrsf37rL04CCPekP
6	Bilinç ve Öznellik	Okuma: Nagel, Thomas (1974). 'What Is It Like to Be a Bat?'. Philosophical Review 83: 435–50, reprinted in Chalmers (ed.) Video: Transformer'lar, LLM'lerin arkasındaki teknoloji: https://youtu.be/wjZofJX0v4M?si=pvah5CyaL2nn8zHq
7	Zor Problem	Okuma: Chalmers, David J. (2002). Consciousness and its place in nature. In Chalmers (ed.) Video: Transformer'larda dikkat mekanizması, adım adım: https://youtu.be/eMlx5fFNoYc?si=CLG8rb_HzQr9XvvM
8	Bilgi Argümanı ve Qualia	Okuma: Jackson, Frank (1982). 'Epiphenomenal Qualia'. The Philosophical Quarterly 32: 127–36, reprinted in Chalmers (ed.) Video: LLM'ler olguları nasıl saklayabilir?: https://youtu.be/9-Jl0dxWQs8?si=MVhGII765FKBfMxn
9	Zihinselliğin İşareti Olarak Yönelimsellik	Okuma: Crane, Tim (2022). 'Intentionality as the Mark of the Mental'. In A. O'Hear (ed.), The Philosophy of Mind, pp. 360–397. Cambridge University Press. Video: Peki yapay zeka görüntüleri ve videoları nasıl çalışıyor?: https://youtu.be/iv-5mZ_9CPY?si=DY_6du8M2queVrxq
10	İçerik Dışsalcılığı	Okuma: Putnam, Hilary (1975). 'The Meaning of Meaning'. Philosophical Papers, Vol. II: Mind, Language, and Reality. Cambridge University Press. Video: Seyrek Otokodlayıcılarla bir Yapay Zekanın Zihnini Okumak: https://youtu.be/krINuMZhJmU?si=091otzqKDUZbFr9F
11	Yapay Zeka ve Zekâ	Okuma: Turing, Alan M. (1950). 'Computing Machinery and Intelligence'. Mind 59: 433–60. Video: Yapay Zekanın Karanlık Maddesi: https://youtu.be/UGO_Ehywuxc?si=X5FhqcRz4RIb52ur
12	Yapay Zekaya Yönelik Zorluklar	Okuma: Searle, John (1983). 'Can Computers Think?', in Minds, Brains, and Science, pp. 28–41. Cambridge, MA: Harvard University Press, reprinted in Chalmers (ed.) Video: Mekanistik Yorumlanabilirlik: Hızlı Bir Tur: https://youtu.be/veT2Vl4vHyU?si=o-joOyruf0N9OstD
13	Mekanistik Yorumlanabilirlik	Okuma: Bereska, L., & Gavves, E. (2024). Mechanistic interpretability for AI safety–a review. arXiv preprint arXiv:2404.14082 Video: LLM'leri Hacklemek: https://youtu.be/Og47aN8aU6M?si=rtLjYiKmqUalP7n-

14	Önermesel Yorumlanabilirlik	Okuma: Chalmers, D. J. (2025). Propositional interpretability in artificial intelligence. Video: Bir Büyük Dil Modeli Bilinçli Olabilir mi?: https://youtu.be/bskf9jyxmMs?si=IcDzHW0P4EIlCVN-

Kadir Has Üniversitesi'nde bir dönem 14 haftadır, 15. ve 16. hafta sınav haftalarıdır.

ZORUNLU ve ÖNERİLEN OKUMALAR

• Turing, Alan M. (1950). 'Computing Machinery and Intelligence'. Mind 59: 433–60.
• Lewis, David (1966). 'An Argument for the Identity Theory'. Journal of Philosophy 63 (1), s. 17–25.
• Putnam, Hilary (1975). 'The Meaning of Meaning'. Philosophical Papers, Vol. II: Mind, Language, and Reality. Cambridge University Press.
• Chalmers, David J. (ed.) (2021). Philosophy of Mind: Classical and Contemporary Readings (İkinci Baskı). New York: Oxford University Press.
• Crane, Tim (2022). 'Intentionality as the Mark of the Mental'. In A. O'Hear (ed.), The Philosophy of Mind, s. 360–397. Cambridge University Press.
• Bereska, L., & Gavves, E. (2024). Mechanistic interpretability for AI safety–a review.
• Harding, C. (2024). Operationalising representation in natural language processing.
• Chalmers, D. J. (2025). Propositional interpretability in artificial intelligence.

DİĞER KAYNAKLAR

- Early Modern Texts: http://www.earlymoderntexts.com/authors/descartes
- Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/
- Writing Philosophy Guide: http://www.jimpryor.net/teaching/guidelines/writing.html
- 3Blue 1Brown and Other Videos:
1. But what is a neural network?: https://youtu.be/aircAruvnKk?si=ggOdpDZgnpCzMDQ0
2. Gradient descent, how neural networks learn: https://youtu.be/IHZwWFHWa-w?si=HCN7wWZ0NeSsooYU
3. Backpropagation, intuitively: https://youtu.be/Ilg3gGewQ5U?si=Kg28tRJ2D8IFR1dK
4. Backpropagation calculus: https://youtu.be/tIeHLnjs5U8?si=Ss-AwfrFQu-He9dv
5. Large Language Models explained briefly: https://youtu.be/LPZh9BOjkQs?si=Wrsf37rL04CCPekP
6. Transformers, the tech behind LLMs: https://youtu.be/wjZofJX0v4M?si=pvah5CyaL2nn8zHq
7. Attention in transformers, step-by-step: https://youtu.be/eMlx5fFNoYc?si=CLG8rb_HzQr9XvvM
8. How might LLMs store facts: https://youtu.be/9-Jl0dxWQs8?si=MVhGII765FKBfMxn
9. But how do AI images and videos actually work?: https://youtu.be/iv-5mZ_9CPY?si=DY_6du8M2queVrxq
10. Reading an AI's Mind with Sparse Autoencoders: https://youtu.be/krINuMZhJmU?si=091otzqKDUZbFr9F
11. The Dark Matter of AI [Mechanistic Interpretability]: https://youtu.be/UGO_Ehywuxc?si=X5FhqcRz4RIb52ur
12. Mechanistic Interpretability: A Whirlwind Tour: https://youtu.be/veT2VI4vHyU?si=o-joOyruf0N9OstD
13. Hacking LLMs: https://youtu.be/Og47aN8aU6M?si=rtLjYiKmqUalP7n-
14. Could a Large Language Model be Conscious?: https://youtu.be/bskf9jyxmMs?si=IcDzHW0P4EIlCVN-

DEĞERLENDİRME SİSTEMİ

Yarıyıl İçi Çalışmaları	**Sayı**	**Katkı Payı (%)**
Katılım	14	28
Ödev	9	72
Dersle İlgili Sınıf Dışı Etkinlikler (okuma, bireysel çalışma vb.)	28	-
Total:	**51**	**100**

İŞ YÜKÜ HESAPLAMASI

Etkinlikler	Sayısı	Süresi (saat)	Toplam İş Yükü (saat)
Ders Saati	14	3	42
Ödev	9	3	27
Dersle İlgili Sınıf Dışı Etkinlikler	28	2	56
Toplam İş Yükü (saat):			**125**

1 AKTS = 25 saatlik iş yükü

PROGRAM YETERLİLİKLERİ (PY) ve ÖĞRENME ÇIKTILARI (ÖÇ) İLİŞKİSİ

#	PY1	PY2	PY3	PY4	PY5	PY6	PY7	PY8	PY9	PY10	PY11	PY12
OC1												
OC2												
OC3												
OC4												
OC5												

Katkı Düzeyi: 1 Düşük, 2 Orta, 3 Yüksek