

コース名	AI・データサイエンスコース				
講座名	—				
科目名	LLM（大規模言語モデル）/生成AIのための深層学習 (拡散モデル編)				
必修・選択	—	単位	—		
概要・目的	<p>近年の画像や映像の生成AIでは性能が非常に向上している。この発展を見るためには、拡散モデルを理解することが必要となり、同時に画像用のトークナイザの進展を理解する必要がある。まずは、拡散モデルの基本モデルであるDDPMモデルや、DDPMのSDE表現を学習します。次に、画像のトークナイザとしての初期のモデルであるVQVAEとこれに対応するCLIPモデルを学習し、画像トークナイザを使用しないアプローチの用いられたDALL-E2やDALL-E3でのアプローチを学習する。さらに、トランスフォーマーを用いた拡散モデルを見る。終盤では、動画拡散モデル、画像編集機能、ビデオ補完技術などの最新技術も見る。</p>				
到達目標	<p>言語のChatGPTの背景にある深層学習技術の理解が進むと、画像や映像ではどのような生成AI技術が開発されているのかという興味が沸く。本講座は拡散モデルを中心として技術の発展にフォーカスを当てて概要を述べる。この講座は、深層学習的視点での理解を深めるための講座で、深層学習に関して基礎的な知識はすでにある人が受講することを期待している。本受講後には、画像や映像の生成AIについての理解が深まり、裏での深層学習の発展についての理解が深まる。</p>				
授業方法	講義	実施形態	遠隔同期 + 遠隔非同期		
評価方法	数回行う小テストによる				
授業項目	1	画像や映像における生成AI技術のイントロ： 機械学習概要、敵対的生成モデルやVAE			
	2	拡散モデル（DDPM）			
	3	連続データ入力に画像用トークナイザを使用するアプローチ：CLIPモデルとVQVAE			
	4	連続データ入力に画像用トークナイザを使用しないアプローチ：DALL-E2やDALL-E3でのアプローチ			
	5	拡散モデルと拡散トランスフォーマー			
	6	拡散モデル（DDPMのSDE表現）			
	7	再キャプショニング技術、Soraでの時空間パッチを生成する技術、インペインティング、画像編集機能など			
	8	動画拡散モデル、3D物理シミュレーション統合技術、ビデオ補完技術など			
	9				
	10				
	11				
	12				
	13				
	14				
	15				
使用教材	講師がスライドなどで提供する。				
特記事項	<p>第2期の日程は 1回目、2回目 非同期ビデオ（10月後半以降開講） 3回目～5回目：1/20（火）→4限、5限、6限 6回目～8回目：1/27（火）→4限、5限、6限</p>				