ТЕМА	ЭЛЕМЕНТ	БАЗОВЫЙ	ПРОДВИНУТЫЙ	ЭКСПЕРТНЫЙ	МАТЕРИАЛЫ
Generative Adversarial Networks	Знания	Классический Goodfellow GAN с бинарной кросс энтропией. Архитектура сверточного генератора / дискриминатора. Какие слои используются в сверточном генераторе для повышения разрешения картинки? Теоретические гарантии сходимости, в чем концептуальная проблема обучения?	Знакомство с архитектурой StyleGAN и вариациями GAN используемых например для задачи SuperResolution. Какие существуют метрики для анализа качества генерации? В чем их недостатки? Различные варианты loss function: hinge loss, relativistic gan.	Понимание устройства StyleGAN 2, какие проблемы решают. Non-leaking аугментации, для чего нужны. Концептуальные слабости GAN (низкая разнообразность). Знакомство с областью контролируемой генерации - GAN-Space и подобное. Инверсия реальных картинок в латентное пространство GAN-а, какие методы существуют?	ARXIV.ORG ARXIV.ORG ARXIV.ORG ARXIV.ORG
	Практическое задание	Обучить GAN на датасете MNIST. Визуализировать генерации.	Обучить SR-GAN на SR датасете по выбору	Взять предобученный StyleGAN, исследовать латентное пространство на направления для editing, визуализировать с разной силой. Определить за что они отвечают. Применить на реальных картинках инвертировав их своим любимым методом.	
	Вопрос для контрольной/ экзамена	Чему равны теоретически оптимальные (идеальные) генератор и дискриминатор в постановке Goodfellow?	Для чего нужен style network в начале StyleGAN?	Почему обучение с non-leaking аугментациями не приводит к коллапсу генераций и проблемам с обучением GAN?	
VAE	Знания	Классический VAE. Вывод формулы ELBO, лосс для обучения, как можно интерпретировать слагаемые? В чем концептуальное отличие от обычного автоэнкодера?	Современные применения VAE для диффузионной генерации. Какие еще лоссы можно использовать для обучения? (LPIPS, perceptual). Что такое VQ-VAE и для чего полезен?	Связь VAE с интерпретируемыми направлениями в латентном пространстве, получение disentangled репрезентаций	ARXIV.ORG ARXIV.ORG ARXIV.ORG ARXIV.ORG
	Практическое задание	Обучить VAE на датасете MNIST. Визуализировать генерации и интерполяцию между латентными кодами двух реальных картинок	Обучить сверточный VAE на датасете FFHQ комбинируя различные лоссы, оценить качество реконструкции	Обучить VAE на датасете MPI3d (https://paperswithcode.com/dataset/mpi3d-disentanglement) и визуализировать как изменение латентной компоненты влияет на реконструкцию	
	Вопрос для контрольной/ экзамена	Вывести ELBO	Почему VAE плохо работают в качестве генеративных моделей, а в основном применяются как обычные автокодировщики? (например как вспомогательный блок при обучении более мощной генеративной модели)	Какие теоретические причины disentanglement при обучении VAE?	
Диффузионные модели	Знания	Score matching и диффузионные модели. Знание базовых алгоритмов генерации диффузионной модельи (p sample, ddim); в чем их разница (детерминистичный и стохастический солвер). Почему на практике в картинных моделях используется unet-подобные архитектуры? Концептуальная сложность (неделная генерация). Что такое classifier free guidance и для чего нужно (трейдофф между качеством и разнообразием)?	Детальное понимание работы ddpm диффузии, знакомство с разными видами расписаний шума (косинусное, линейное). Продвинутые алгоритмы сэмплирования (dpm++, unipc). Понимание концепта латентной диффузии и почему она так популярна. Методы генерации картинок высокого разрешения не используя латентное пространство. Знание архитектур моделей DiT. Модификации unet для генерации видео.	Знание EDM и EDM2 подходов и какие проблемы они решают. Rectified Flow	ВСЕ СТАТЬИ С YANG-SONG.NET ARXIV.ORG ARXIV.ORG ARXIV.ORG
	Практическое задание	Обучить диффузионную модель (DDPM с линейным скедулером) и обучить на двумерном датасете из смеси гауссиан расположенных по сетке (центры гауссиан находятся в узлах решетки), например размера 4x4. Визуализировать генерации моделей. для диффузионной модели реализовать и применить алгоритм p-sample. Сравнить результаты различных моделей, качественно оценить результаты.	Обучить картиночную class-conditional диффузионную модель (DDPM с линейным скедулером) на датасете CIFAR. Реализовать CFG. Провести исследование зависимости качества обучение модели (метрики FID, Precision & Recall) в зависимости от значения CFG. Для генерации воспользоваться методом dpm++. Сформулировать выводы о том как меняются метрики, и что мы получаем/теряем на практике, беря высокие значения CFG.	Реализовать EDM алгоритм. Обучить диффузионные class-conditional модели на CIFAR со стандартным линейным скедулером / EDM алгоритмом. Сравнить результаты.	
	Вопрос для контрольной/ экзамена	Что такое score function в диффузионных моделях и каким образом мы учим ее на практике? Как знание ее помогает генерировать данные?	Каким образом текстовая информация подается в text-to-image модели? Какие существуют опции для генерации картинок высокого разрешения? (1024+).	Какие проблемы решают EDM/EDM-2 алгоритмы? Что такое rectified flow подход, для чего он нужен? На практике, какое расписание шума стоит использовать? (анализ существующих например из статьи SD3).	
Практические применения диффузионных моделей	Знания	Как работает инпейнтинг через диффузию. Метод SDEdit. Что такое LoRA и для чего используется, какие есть вариации. ControlNet. Для чего нужна текстовая инверсия и что такое персонализация. Примеры методов.	Что такое IP adapter? Как можно добиться возможности кондишенинга модели на разные виды данных - маска, canny, непрерывные величины, bbox. Какие есть методы guidance кроме стандартного CGF? (PAG, SAG). Какова их стоимость в терминах форвард пассов модели	Какие подходы есть к задаче ImageBlending? (Смержить две картинки) Какие подходы есть к задаче FaceRemix? (Переписовать лицо по данному промпту). Remix картинок по текстовому промту.	STABLE-DIFFUSION-ART.COM ARXIV.ORG WWW.JUSTINPINKNEY.COM ARXIV.ORG
	Практическое задание	Взять StableDiffusion 1.4 и обучить LoRA на датасете по выбору (для примера датасет покемонов). Для кашшенинга картинок можно взять VLM, например LLaVA 1.5. Оценить качество глазами, подобрать оптимальное значение ранга в LoRA.	Взять StableDiffusion 1.4 и дообучить с кондишеингом на bbox объекта. Потенеировать с обуславливанием на bounding box.	Обучить модель реализующую ImageBlending поверх SD 1.4. Визуализировать блендинг двух картинок	
	Вопрос для контрольной/ экзамена	Какие параметры можно варьировать в LoRA? Как варьируется "сила" LoRA?	Каким слоем можно подать непрерывную информацию в модель? Например, обуславливание на aspect ratio картинки	Как можно варьировать сколько "высокоуровневых фичей" сохраняем из картинки при remix?	
Диффузионные модели: теория	Знания	Score matching, score function - теоретические основы и как реализуются на практике. Динамика Ланжевина. Знание точки зрения на диффузию через стохастическое дифференциальное уравнение. Уравнение Фоккера-Планка. Probability flow ODE.	Точка зрения на DDPM диффузию (процесс Орнстейна-Уленбека) как на градиентный спуск в пространстве плотностей. Скорость сходимости через W_2 расстояние.	Диффузия не через винеровский процесс, а произвольные деформации картинки (блуры).	ARXIV.ORG CEDRICVILLANI.ORG
	Практическое задание	Доказать что denoising score matching в оптимуме дает score function приближенного распределения.	Даны две гауссианы N(0, 1) и N(a, b). Чему равно W_2^2 расстояние между ними. Как будет убывать расстояние если мы начнем диффузионный процесс стартуя с N(a, b)	Визуализировать процесс размытия картинки гауссианским ядром при варьировании размера ядра.	
	Вопрос для контрольной/ экзамена	Как выглядела бы DDPM диффузия в случае когда терминальное распределение не N(0, 1) а имело бы плотность Z exp(-||x||^4)	Выписать W_2^2 расстояние между двумя произвольными гауссианами в R^n	Ограничения данных методов, почему их не применяют в практических сценариях.	
Оценка качества генеративных картиночных моделей	Знания	Оценка качества через Inception Score, FID, KID, MMD. Недостатки данных метрик. Как считается и какой физический смысл метрики FID. Какие еще модели вместо Inception V3 можно использовать?	Как замеряется качество text-to-image генерации, CLIP/FID кривые? Баланс между понижем текста и визуальным качеством генераций	Оценка качества с помощью side-by-side разметки, использование людей-аксессоров. Какие сервисы есть (Яндекс Толока, Amazon mturk)? Использование полученной разметки в качестве сигнала для reinforced learning with human feedback.	ARXIV.ORG ARXIV.ORG WWW.MTURK.COM
	Практическое задание	Взять два непересекающихся подмножества CIFAR размера N, построить график FID между данными выборками при возрастании N. Так же оценить стандартное отклонение полученных значений.	Посчитать clip/fid кривую для SD 1.4 , варьируя значения CFG	Сделать прототип инструмента оценки качества, сравнить SD 1.4 и SDXL. За основу можно взять публичный инструмент https://labelstud.io/ или же библиотеки для быстрого прототипирования (https://streamlit.io/)	
	Вопрос для контрольной/ экзамена	Привести пример игрушечных распределений с небольшим значением FID но радикально отличающихся	Каким типичным образом меняются значения FID/CLIP при варьировании CFG в диффузионной text-to-image модели?	По каким аспектам можно сравнивать 2 картиночные генерации по текстовому входу? (Релевантность, эстетичность, и тд).	
Ускорение диффузионных моделей	Знания	Медленная скорость генерации как недостаток диффузионных моделей. Методы ускорения - guidance distillation, consistency model. Использование LORA для обучения дистиллированной модели.	Adversarial Diffusion Distillation, Imagine Flash методы. Backward distillation.	Distribution Matching Distillation, Improved Distribution Matching Distillation.	ARXIV.ORG ARXIV.ORG ARXIV.ORG ARXIV.ORG ARXIV.ORG
	Практическое задание	Реализовать guidance distillation для SD 1.4. Сравнить метрики генерации для исходной модели с дистиллированной.	Реализовать ADD для SD 1.4. Сравнить метрики генерации для исходной модели с дистиллированной.	На базовой модели SD 1.4 провести сравнение современных методов дистилляции: Guidance Distillation, Latent Consistency Model, ADD, DMD, DMD2.	
	Вопрос для контрольной/ экзамена	Обсудить какая модель потенциально лучше дистиллируется - обученная с помощью RF постановки диффузии или классической DDPM.	Какие есть варианты для архитектуры дискриминатора? Как можно использовать предобученную диффузионную модель для этого?		
Диффузионные модели для генерации видео	Знания	Псевдо-видео модели. Deforum; возможность анимации через outpainting (сдвиг картинки + дорисовка).	End-to-end генерация видео. Двухстадийный пайплайн: модель для ключевых кадров + модель интерполяции. Temporal attention, RoPE. Align Your Latents	SOTA методы генерации: Moviegen (text to image + video); CogVideo. Использование темпорального VAE как альтернатива модели интерполяции кадров.	GITHUB.COM VK.COM ARXIV.ORG ARXIV.ORG GITHUB.COM
	Практическое задание	Взять любую модель для аутпейнтинга SD по выбору, наобавить баннеров по выбору и реализовать их анимирование через outpainting. Сделать выводы о качестве.	Взять предобученную диффузионную модель (SDXL, Flux). Дообучить модель генерации ключевых кадров на датасете WebVid. Провести исследование и аблейшен того как подается темпоральная информация в модель (темпоральные аттеншены, использование RoPE)	Дообучить VAE от диффузионной модели по выбору (SDXL,Flux) до темпорального на видеодатасете по выбору (например WebVid). Затем дотюнить на нем выбранную диффузионную модель. Проанализировать качество полученное модели	
	Вопрос для контрольной/ экзамена	Какие преобразования используются в deforum?	Почему выгоднее использовать две модели (ключевые кадры + интерполяция), а не учить end-to-end модель?	Чем парадигма темпоральных вае лучше чем модели интерполяции?	
Особенности text-to-image генерации	Знания	Основные представители семейства - Stable Diffusion, DALLE-3, Imagen. Две парадигмы (латентная и каскадная). В чем их отличия и какая более вычислительно эффективна? Какие модели используются в каскаде, как получаются латентные коды в латентном подходе? Как модель получает информацию о тексте?	Обработка кондишенинга в виде текстовых эмбеддингов. Какие существуют способы его преобразовать (LN, PerceiverResampler). CrossAttention между токенами? Как можно улучшить реагирование модели на текст, что такое негативный промт, null embedding?	Существующие техники улучшения качества обучения (conditioning на aspect ratio картинки как в SDXL., диффузионные трансформеры как в SD3/FLUX). Подготовка данных для обучения с помощью моделей-капшенеров (LLaVA).	ARXIV.ORG OPENAI.COM IMAGEN.RESEARCH.GOOGLE GITHUB.COM STABILITY.AI
	Практическое задание	Обучить простую диффузионную text2image модель на датасете COCO в разрешении 64x64.	Изучить влияние негативного промта на генерации SD 1.4	Выбрать капшенер по вкусу и сравнить как вариации подходи (Describe an Image...) влияет на качество файнтюна SD 1.4 на датасете по выбору.	
	Вопрос для контрольной/ экзамена	Как получить текстовые эмбеддинги для подачи в модель?	Какие есть слои которые помогают модели учитывать текстовую информации при генерации?	Какие свойства капшенов хороши для обучения модели? Как можно описать их разницу от естественных текстов из интернета?	