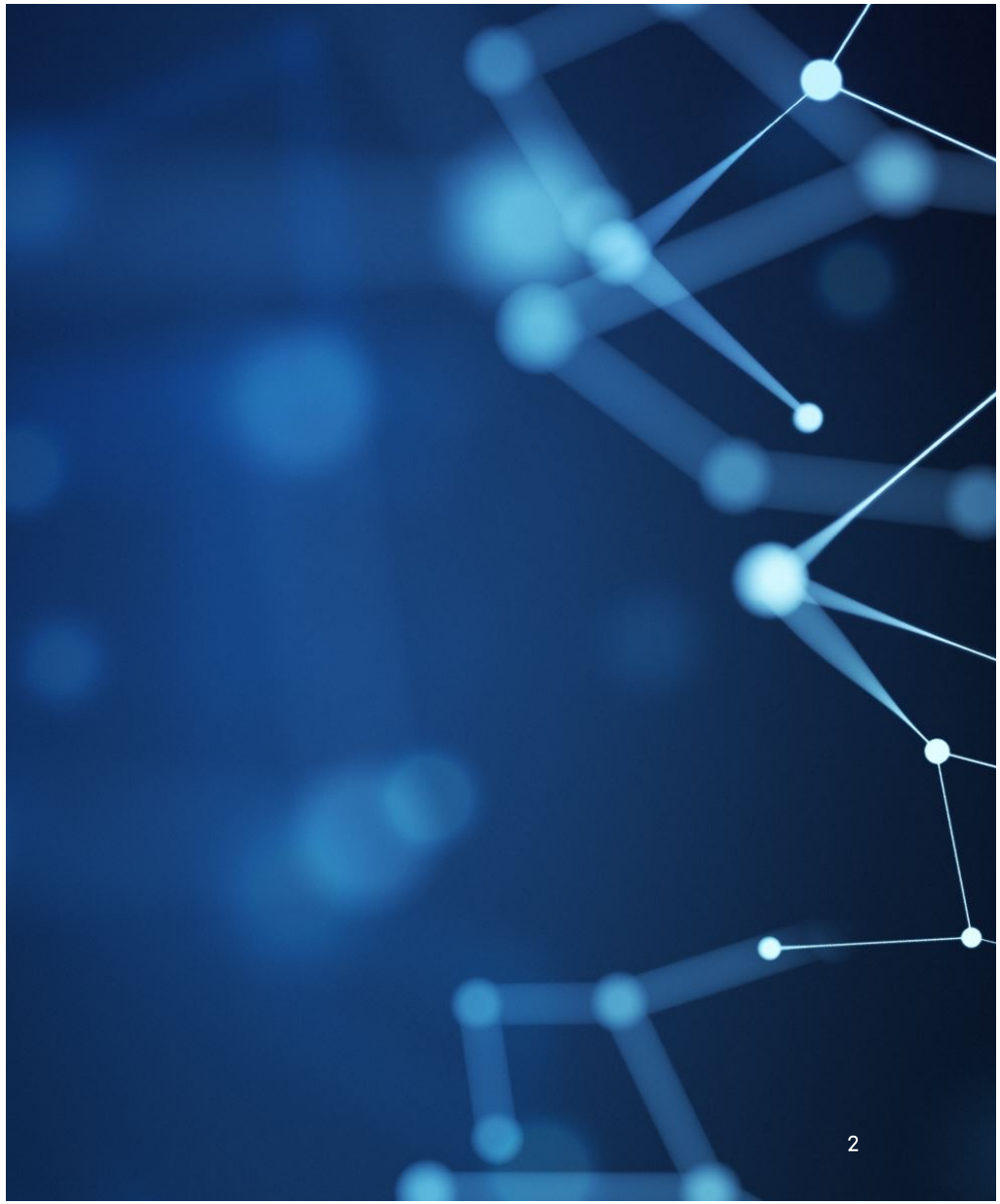
Text 2 Video

VIDEO GENERIERUNG MIT
BASISMODELLEN DER KI

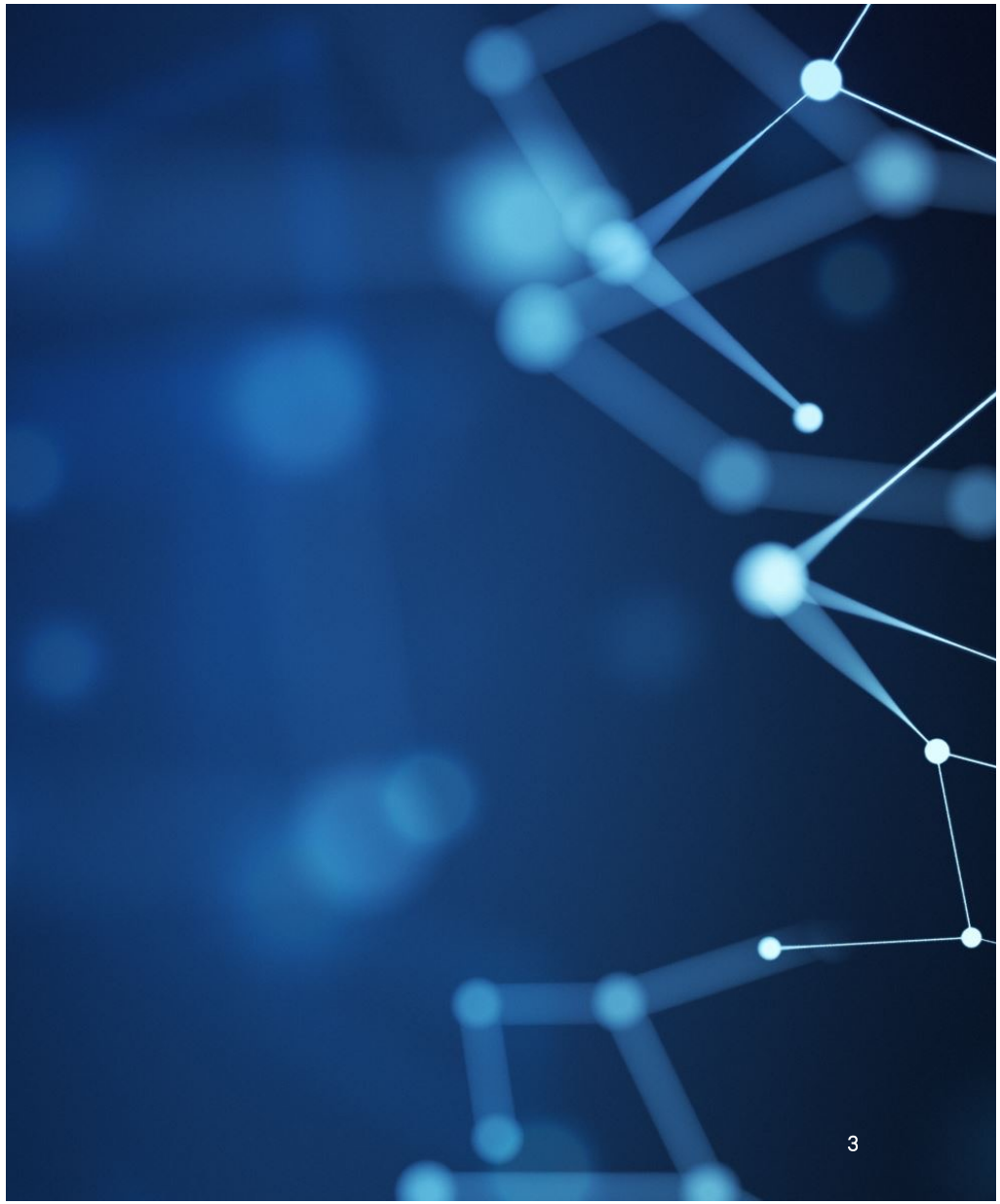
Gliederung

1. WAS IST T2V?
2. FUNKTIONSWEISE
3. ARCHITEKTURÜBERSICHT
4. TOOLS UND BEISPIELE
5. HERAUSFORDERUNGEN
6. FAZIT



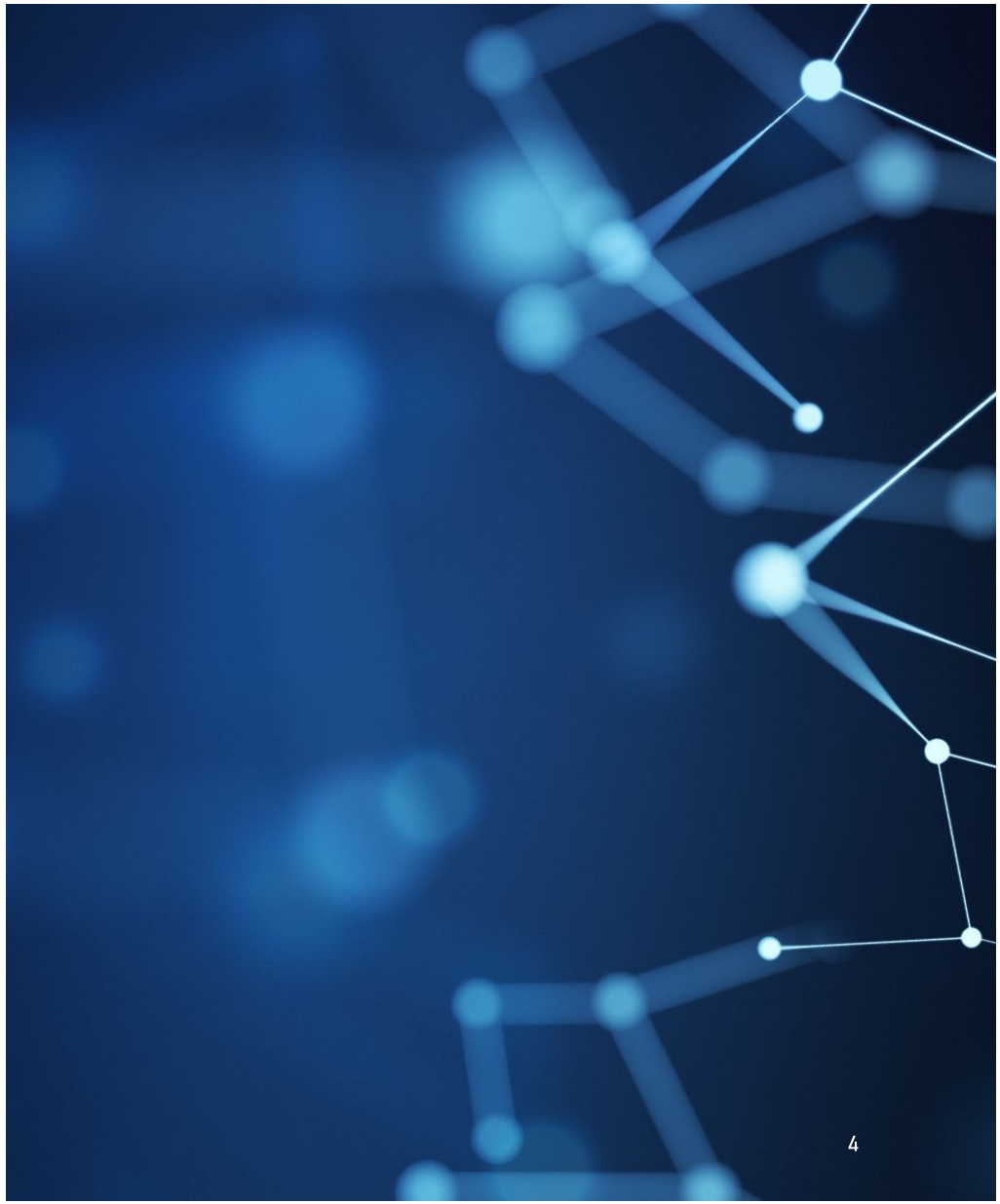
Was ist T2V

TEXT-TO-VIDEO(T2V) IST EINE TECHNOLOGIE, DIE AUS EINEM TEXT EIN VIDEO ERZEUGT



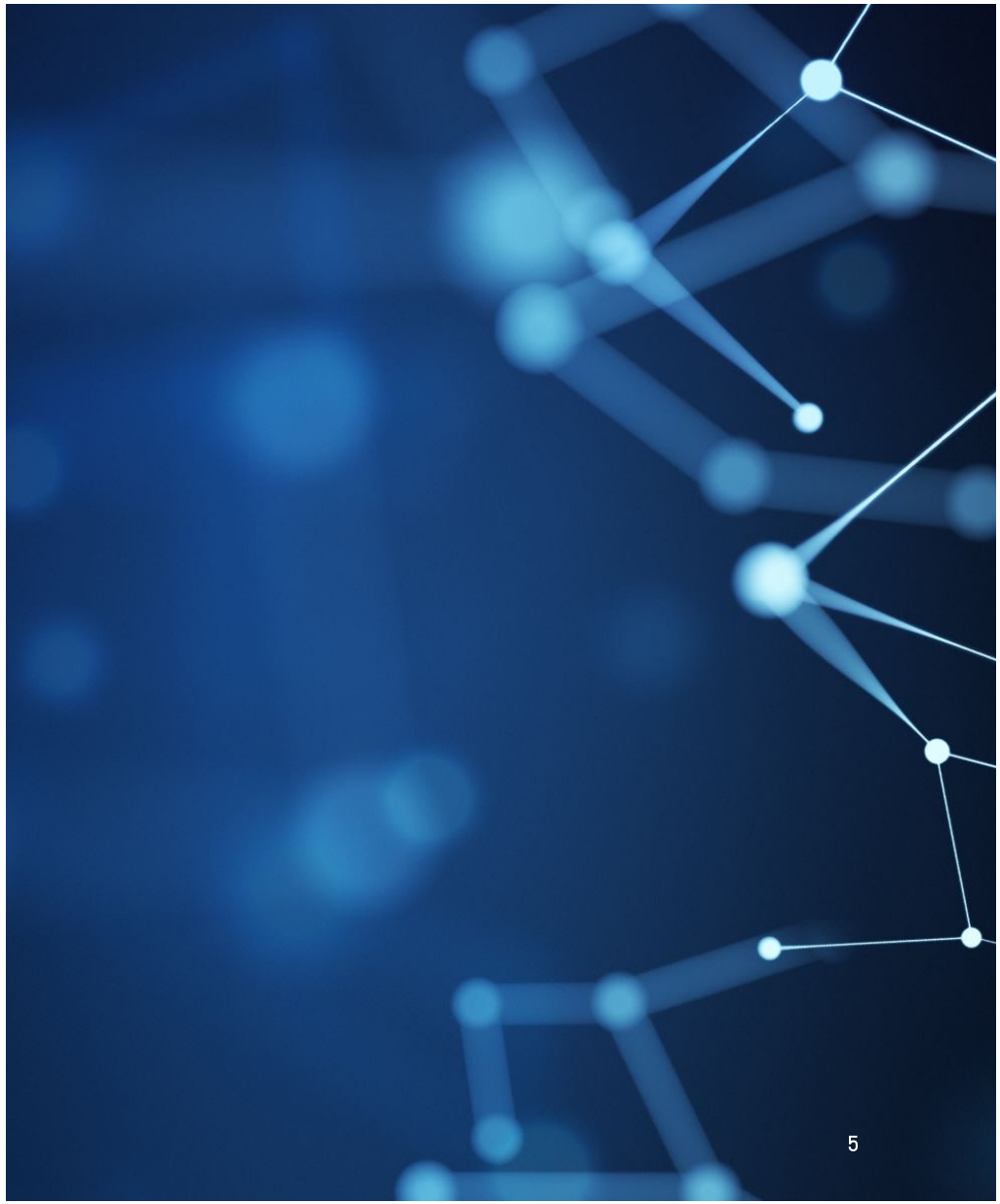
Was bringt T2V

- REDUZIERT PRODUKTIONSZEIT & KOSTEN (FILM, WERBUNG, BILDUNG, SOCIAL MEDIA)
- VIDEOPRODUKTION OHNE KAMERA UND SCHAUSPIELER



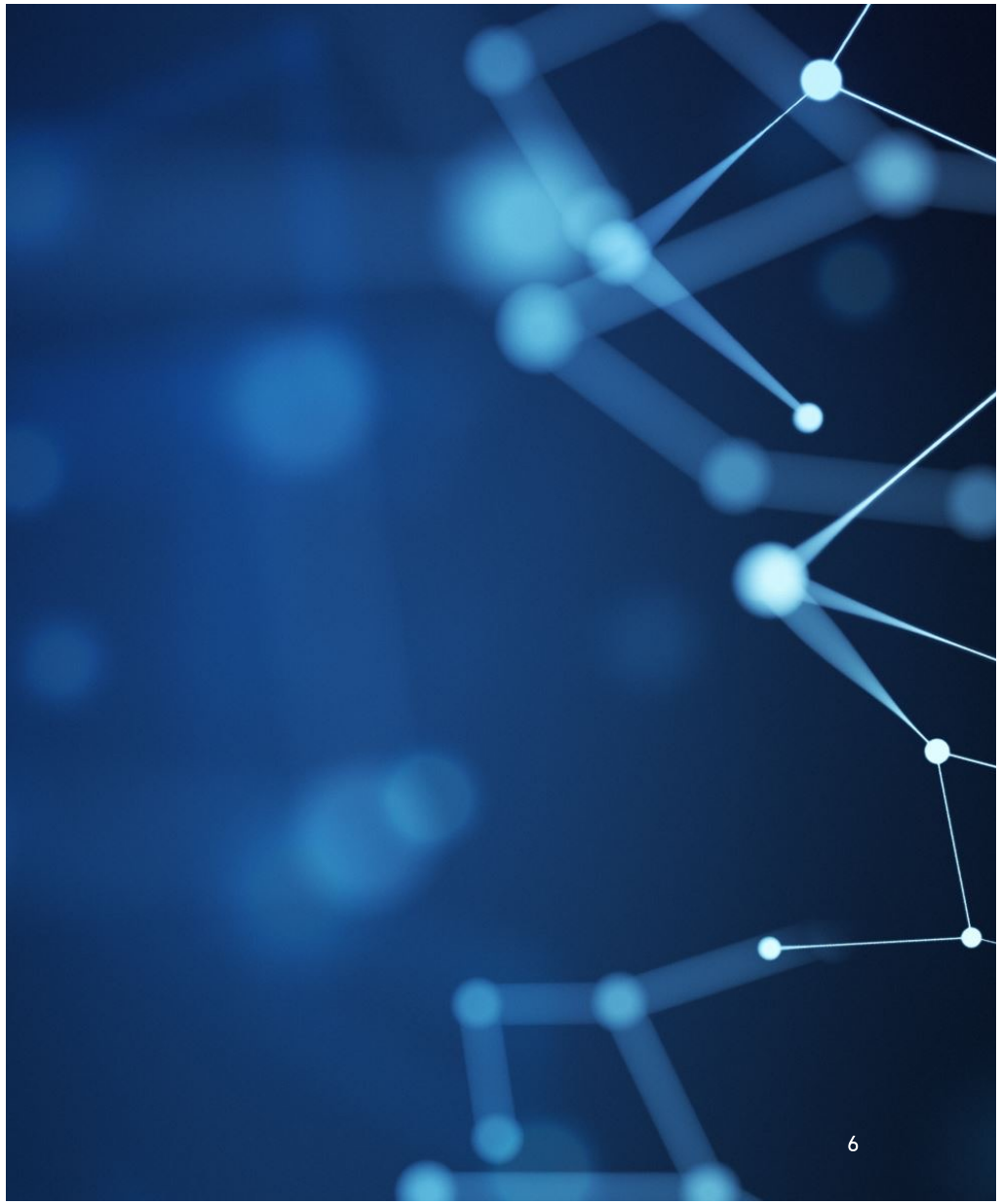
Wie funktioniert T2V?

1. TEXT DECODEN
   KI LIEST TEXT UND ERKENNT WAS VORKOMMT
2. NOISE GENERIEREN.
   KI STARTET MIT ZUFALLSVERRAUSCHTEM BILD
3. SCHRITTWEISE DENOISING
   DIFFUSSIONSMODELL ENTFERNT NOISE IN VIELEN KLEINEN SCHRITTEN
4. POSTPROCESSING
   FARBKORREKTUR, UPSCALING



T2V Architekturen

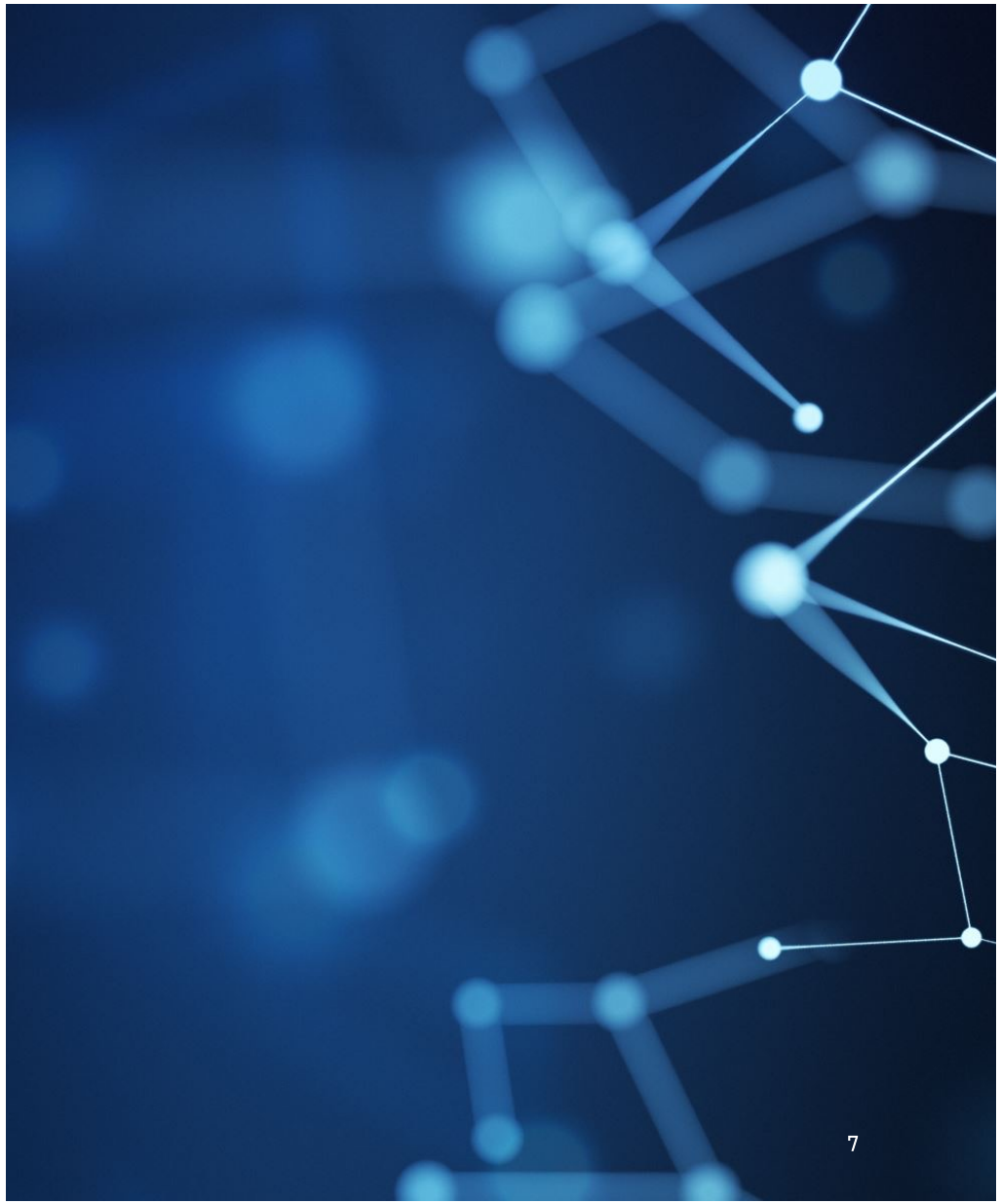
- VAE (VARIATIONAL AUTOENCODER)
- GAN (GENERATIVE ADVERSARIAL NETWORK)
- DM (DIFFUSIONMODELL)
- DIT (DIFFUSION TRANSFORMER)



DiT (Diffusion Transformer

GRUNDIDEE:

- DIFFUSION KÜMMERT SICH UM DIE BILDGENERIERUNG
- TRANSFORMER VERSTEHT ABHÄNGIGKEITEN ZWISCHEN TEXT UND ZEIT (KOHÄRENZ & PHYSIK
- KOMBINATION ERMÖGLICHT LÄNGERE UND REALISTISCHERE VIDEOS



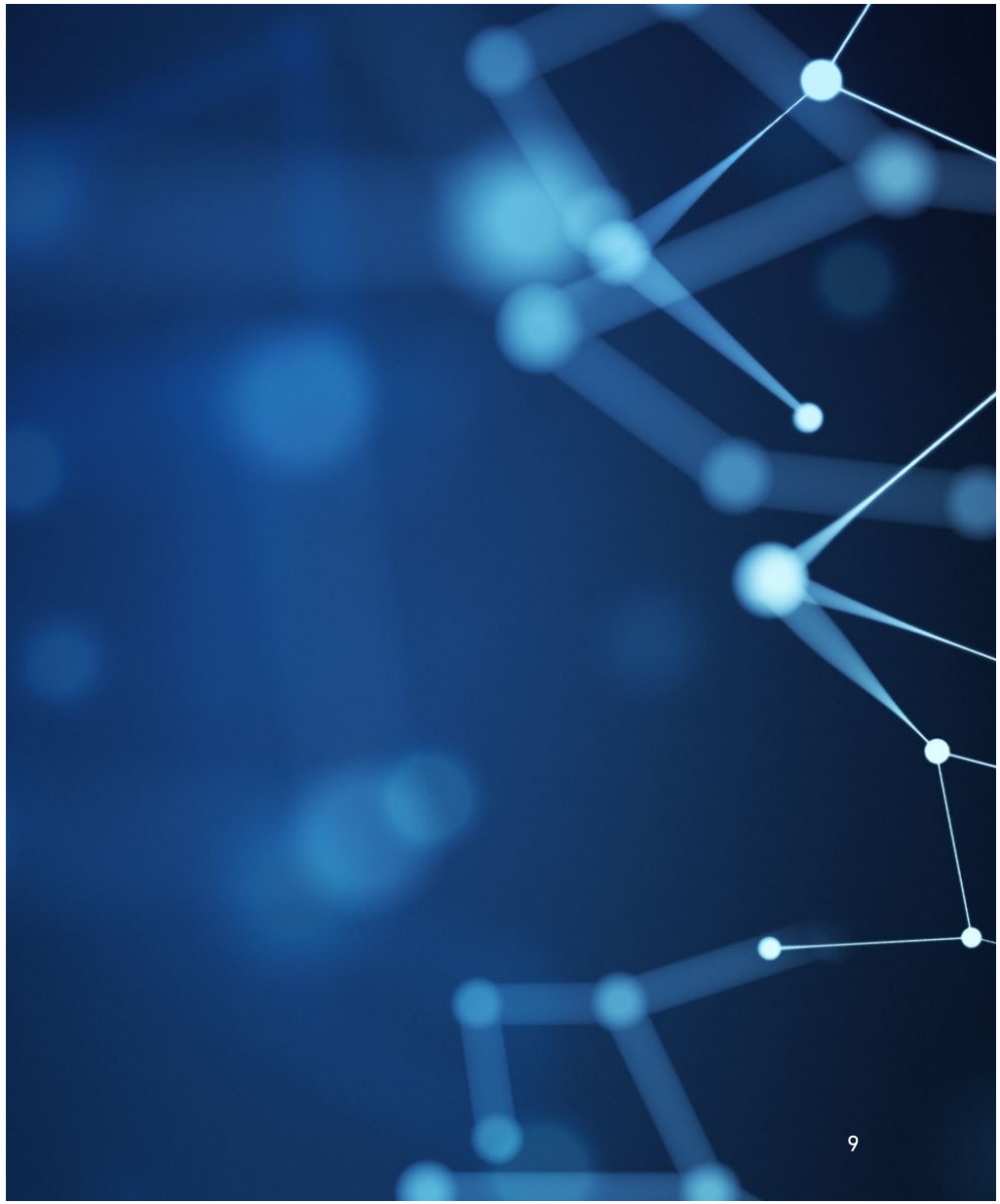
DiT (Diffusion Transformer

WIE FUNKTIONIERTS ?

- ES WIRD EIN TEXT EINGEGEBEN
- DER TEXT WIRD DURCH DEN TRANSFORMER VERSTANDEN
- IM LATENTEN RAUM WIRD NACH PASSENDEN UND VORHANDENEN INFORMATIONEN GESUCHT
- DAS VIDEO WIRD DENOISED
- DAS FERTIGE VIDEO WIRD ERSTELLT

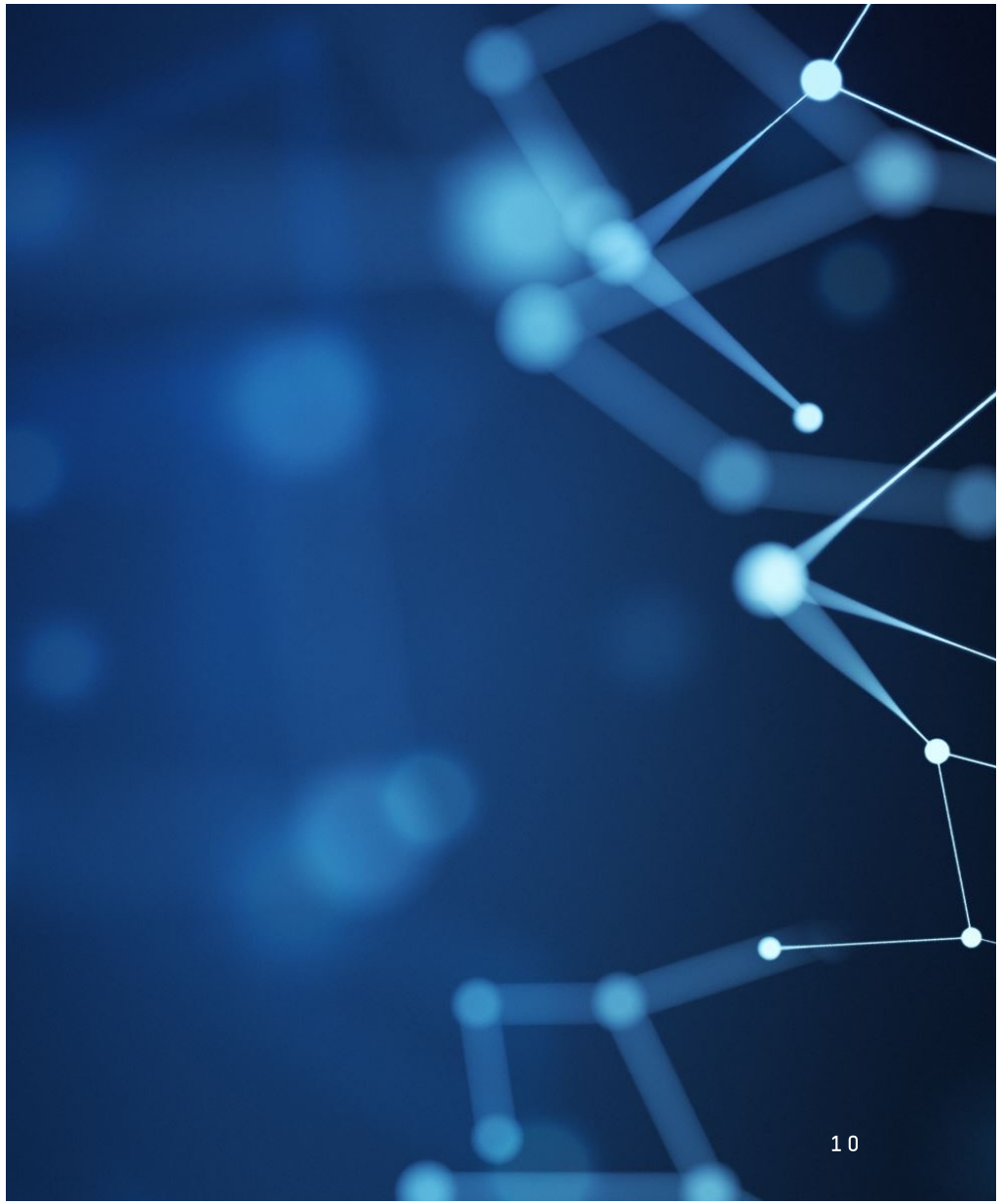
Eigenschaften von DiT

- SPATIOTEMPORAL UNDERSTANDING (RAUM + ZEIT)
- CROSS-ATTENTION MECHANISMUS (TEXT & VIDEO AUFEINANDER ABGEGLICHEN)
- REALISTISCHE BEWEGUNGEN
- SKALIERBARKEIT (MEHR DATEN + RECHENLEISTUNG)



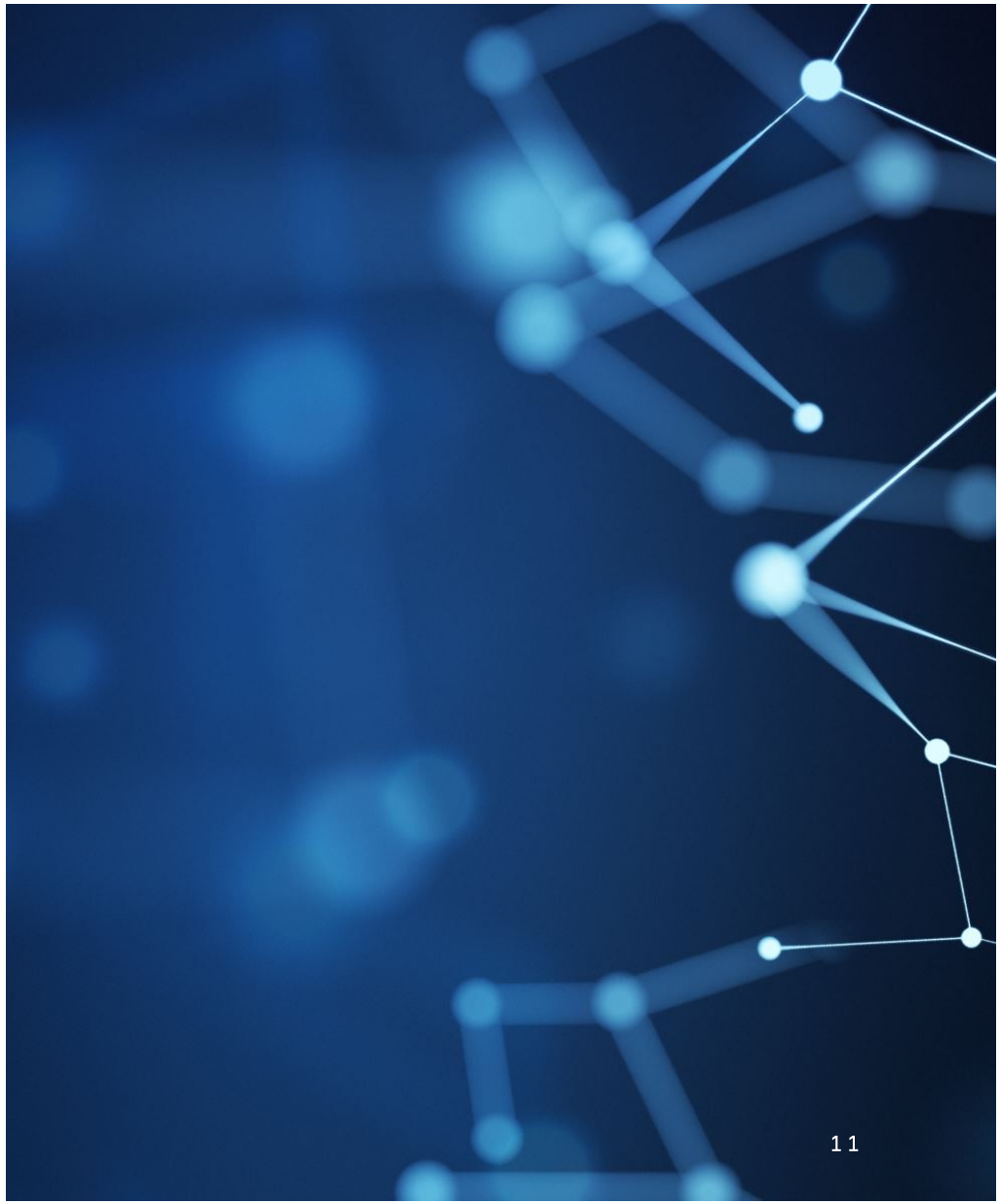
Tools und Beispiele

- BEISPIELPROMPT:
  STUDENT PRÄSENTIERT EINE KI-PRÄSENTATION
- SORA
- PIKA
- RUNAWAY



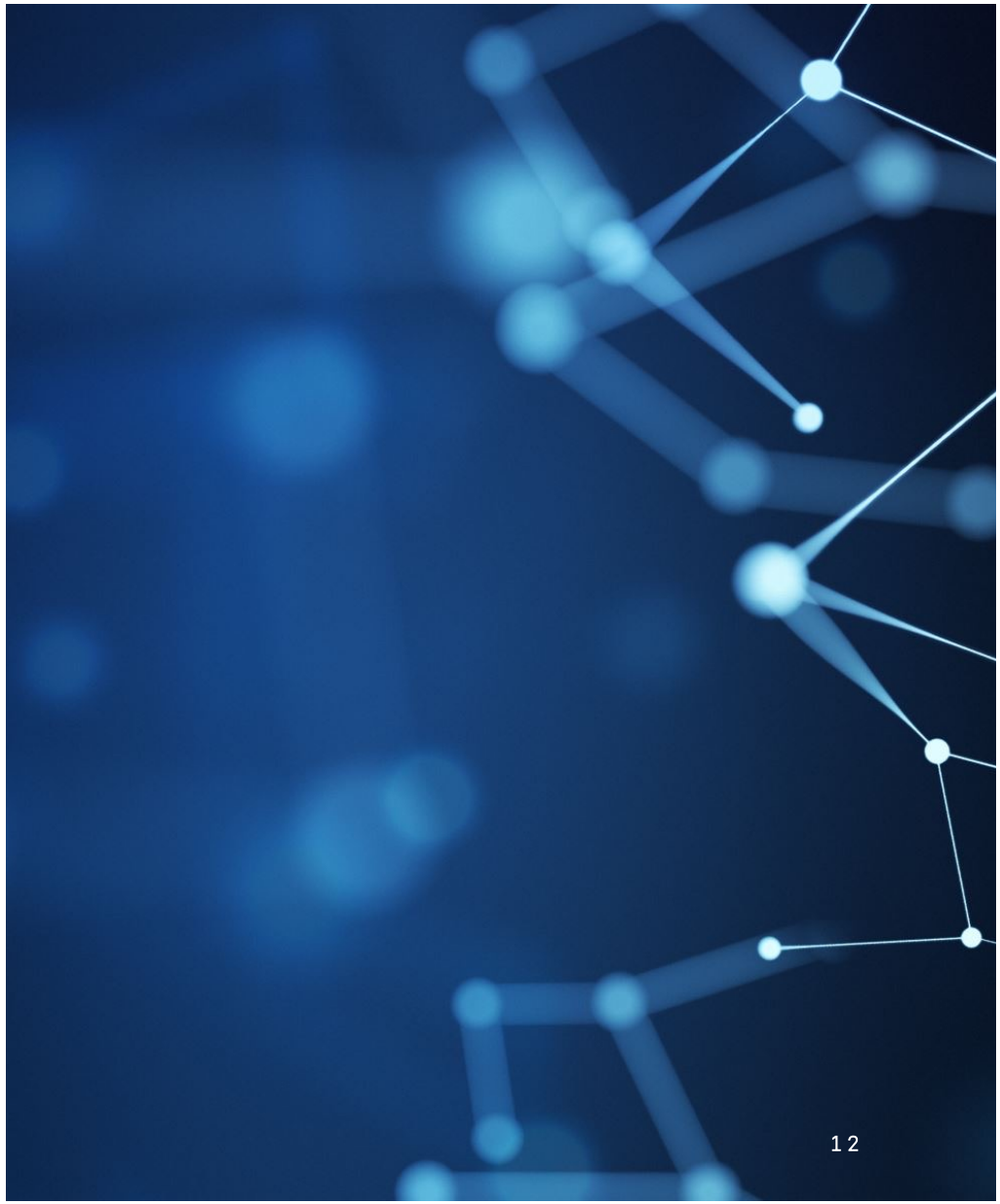
Herausforderungen

- VERSTÄNDNIS KOMPLEXER PROMPTS
- KONSISTENZ ZWISCHEN FRAMES
- PHYSIKALISCHER REALISMUS
- RECHENAUFWAND & ENERGIEVERBRAUCH
- DATEN & URHEBERRECHT
- MISSBRAUCH & DEEPFAKES



Fazit

- KI KANN HEUTE AUS TEXT REALISTISCHE VIDEOS ERZEUGEN
- DIFFUSION TRANSFORMERS SIND DER AKTUELLE STAND DER TECHNIK
- NOCH GROßE HERAUSFORDERUNGEN WIE PHYSIK, ENERGIE UND ETHIK
- T2V HAT EIN GROßES POTENTIAL FÜR FILM, BILDUNG UND WERBUNG



Quellen

- https://www.youtube.com/watch?v=iv-5mz_9cpy
- https://www.youtube.com/watch?v=szoraj4i-sa
- https://arxiv.org/pdf/2510.04999
- https://www.youtube.com/watch?v=iv-5mz_9cpy&t=527s
- CHAT GPT (STAND 10.11.2025)
- https://www.youtube.com/watch?v=zxirugocn9s
- https://www.youtube.com/watch?v=dfciadfn2d4
- https://www.youtube.com/watch?v=0_caoa5fmhe
- 3blue1brown dl playlist