

Die Lange Nacht der Wissenschaften

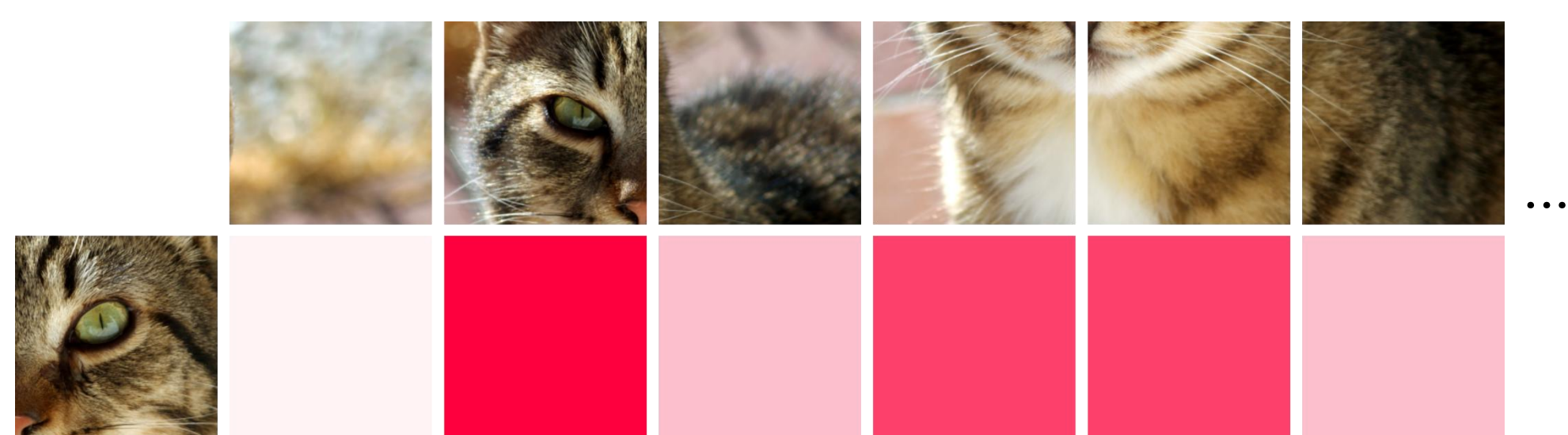
Einblicke in KI-Bildmodelle - wie moderne Vision-Transformer Bilder verstehen

KI ist überall: in den Medien, im Alltag und in der Wissenschaft. Aber was läuft im Hintergrund, wie verstehen grafische KI-Modelle Bilder? Hier zeigen wir Einblicke in das **Innenleben** von sog. Vision-Transformern – den Arbeitstieren in den Bereichen Klassifikation, Bild-zu-Text und Bilderzeugung.

Bild Klassifikation

1. Bild skalieren und zuschneiden
2. Tokenisierung: Bild aufteilen und als Zahlen (*Token*) repräsentieren
3. Modell weist jedem Token einen *Aufmerksamkeitswert* zu
 - Mehrere Aufmerksamkeitsköpfe (Heads) parallel
 - Mehrere Blöcke hintereinander (Layer)
4. Klassifikation: sog. Multi-Layer-Perceptron (MLP) sammelt alle Aufmerksamkeitswerte des letzten Layer und aller Heads und weist diese einer Klasse zu

Aufmerksamkeit – Self-Attention



Wie stark hängt ein Ausschnitt mit jedem anderen zusammen (Korrelation)?

Einblicke zur Aufmerksamkeit:

1. Aufmerksamkeitsköpfe lernen, Muster zu erkennen
 - Frühe Layer fokussieren sich auf oberflächliche Muster
 - Späte Layer erkennen abstrakte Muster
2. Parallele Köpfe erkennen unterschiedliche Muster, z.B. Hintergrund, Vordergrund, Geometrien, Farben...

Wie stark tragen Bildbereiche zur Vorhersage bei:

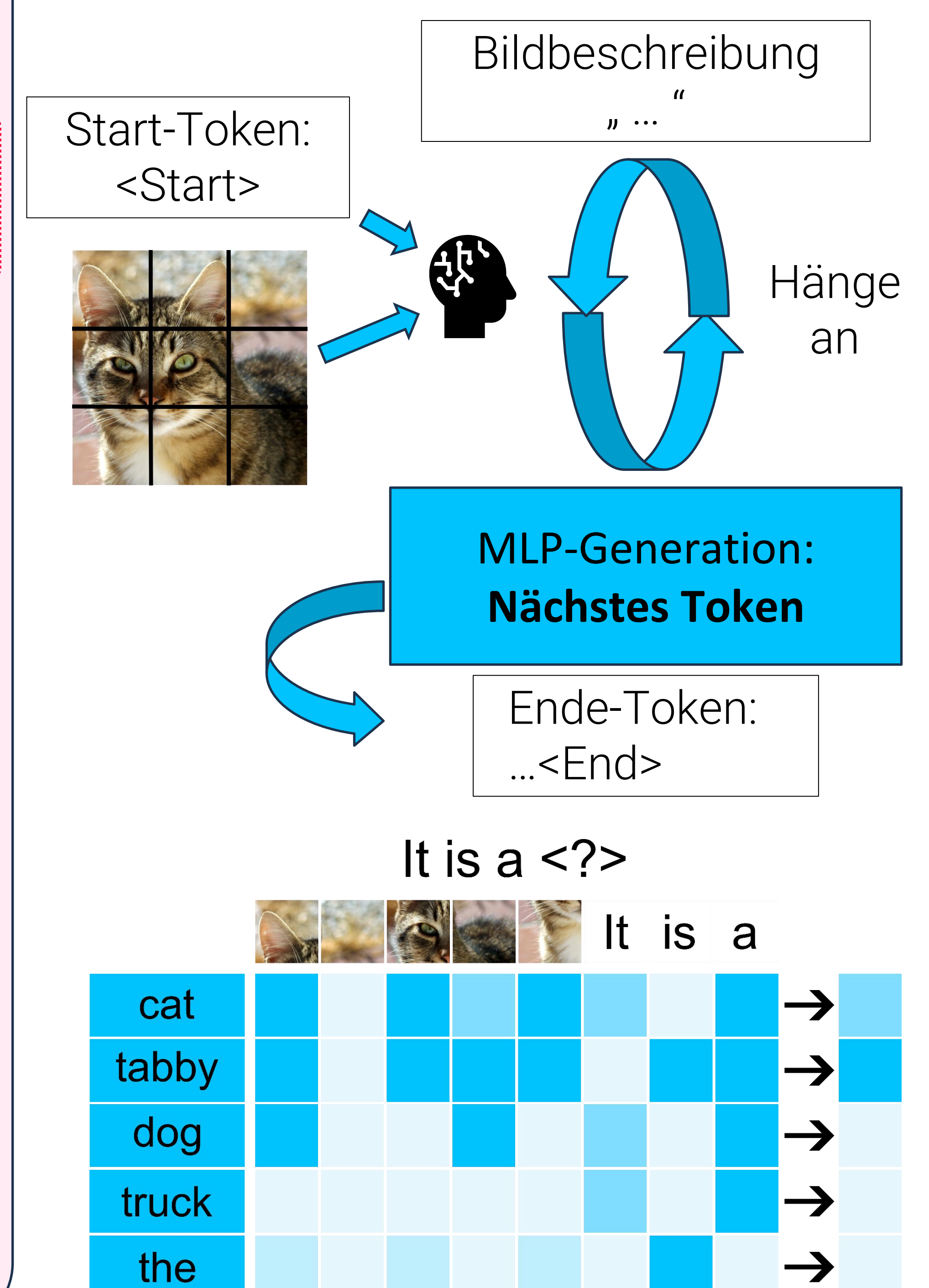
1. Vergleich mit komplett schwarzem Bild – welche Bildbereiche tragen am stärksten zur Vorhersage bei?
2. Identifikation der Bildabschnitte bzgl. Bild-zu-Text-Token
3. Überraschende Sensitivitäten, z. B.: Das Fell einer Katze trägt zur Erkennung von Gras im Hintergrund bei

Lassen Sie sich die Einblicke live an unserem Stand zeigen!



Bild-zu-Text

1. Übergebe zusätzlich zum Bild ein Start-Token
2. ...
3. MLP generiert das wahrscheinlichste nächste Token
4. Wiederhole bis MLP das Ende-Token generiert



Bezug zum MPL

In der Gruppe von Prof. Florian Marquardt untersuchen wir, wie moderne KI im wissenschaftlichen Prozess von morgen verwendet werden könnte. Kann KI selbständig die wissenschaftliche Methode anwenden und neue Phänomene entdecken? Welche Eigenschaften muss KI besitzen, um nachvollziehbare Vorhersagen und Schlussfolgerungen zu ziehen? Wo liegen die Grenzen der heute verfügbaren Modelle?



MAX-PLANCK-ZENTRUM
FÜR PHYSIK UND MEDIZIN



MAX-PLANCK-INSTITUT
FÜR DIE PHYSIK DES LICHTS