

A complex network diagram with various colored nodes (circles) and connecting lines, set against a light background. The nodes are in shades of blue, red, orange, and purple, and are connected by thin grey lines. Some nodes are larger and more prominent than others.

# Cognitive AR

Endbericht | Call 15 | Projekt ID 5121

Lizenz CC-BY-SA-4.0

# Inhalt

1 Einleitung.....	3
2 Projektbeschreibung.....	3
3 Verlauf der Arbeitspakete.....	4
3.1 Arbeitspaket 1 - AP1: Detailplanung und Formales am Projektstart.....	4
3.2 Arbeitspaket 2 - AP2: Design & Technologie Evaluation.....	4
3.3 Arbeitspaket 3 - AP3: Computer Vision / Artificial Intelligence Komponenten.....	4
3.4 Arbeitspaket 4 - AP4: Plattform Implementierung.....	4
3.5 Arbeitspaket 5 - AP5: Augmented Reality Apps.....	5
3.6 Arbeitspaket 6 - AP6: Demo-Deployment.....	5
3.7 Arbeitspaket 7 - AP7: Dissemination.....	6
3.8 Arbeitspaket 8 - AP8: Dokumentation und Formales am Projektende.....	6
4 Umsetzung Förderauflagen.....	6
5 Liste Projektendergebnisse.....	6
6 Verwertung der Projektergebnisse in der Praxis.....	7
7 Öffentlichkeitsarbeit/ Vernetzung.....	8
8 Eigene Projektwebsite.....	9
9 Geplante Aktivitäten nach netidee-Projektende.....	9
10 Anregungen für Weiterentwicklungen durch Dritte.....	9

## 1 Einleitung

Der folgende Bericht erläutert die Fortschritte von „Cognitive AR“, das erste Projekt der größeren Cognitive XR Initiative. Das Ziel von Cognitive AR ist der Aufbau einer Software Plattform und Client-Server Struktur, für das Integrieren von Augmented Reality Brillen mit einer dezentralen Edge Computing Infrastruktur. Zuerst möchten wir uns für die Unterstützung der Netidee bedanken die es uns ermöglicht hat dieses spannende Projekt und in unseren Augen zukunftsweisende Thema zu bearbeiten. Die Netidee Förderung ermöglicht uns sowohl die Open Source Entwicklung der Software, als auch das Publizieren unsere Ergebnisse in Blogs und auf Konferenzen. Wir lagen im Projektplan gesamt etwa vier Monate zurück, und haben deshalb um eine Projekt- verlängerung angesucht, die Sie uns dankenswerterweise genehmigt haben. Wir sind mit dem Ausgang des Projekts insgesamt zufrieden. Bei etwa 40% der Ergebnisse haben wir den im Antrag definierten ausgezeichneten Erfolg erreicht, und bei 60% den minimalen Erfolg.

## 2 Projektbeschreibung

Wir sind zukunftsorientierte Forscher\_innen und Ingenieure in den Bereichen Cloud Computing, Internet der Dinge, AI, und Virtual/Augmented Reality (AR). Wir sind überzeugt dass Cognitive Assistance Anwendungen mit wearable AR einen immensen Mehrwert für die digitalisierte urbane Gesellschaft bieten.

Entwickler\_innen von Apps für AR wearables, z.B. Smartglasses, können mit unserer Plattform Anwendungen mit Echtzeitanforderungen umsetzen, die rechenaufwendige AI Services, oder Sensordaten aus der unmittelbaren Umgebung im urbanen Raum benötigen. Solche Cognitive Assistance Anwendungen können z.B. Bürger\_innen beim Radeln durch die Stadt unterstützen, oder Sehbehinderten beim Einkaufen helfen.

AR wearables haben i.d.R. sehr beschränkte Rechen- und Speicherkapazität. Funktionalität wird daher häufig durch Cloud-basierte services implementiert, was jedoch für Echtzeitanwendungen (z.B. Objekterkennung für ein Kollisionswarnsystem), wegen der hohen Latenz nicht ausreicht. Um diese Anwendungen zu ermöglichen benötigt man eine stadtweite Edge Computing Infrastruktur und Laufzeitumgebung, die gleichzeitig den einfachen Zugriff auf Sensordaten von z.B. Smart Lampposts in Echtzeit erlaubt.

Ziel unseres Projektes war es eine Plattform zu entwickeln, die den skalierbaren Betrieb von Services für Cognitive Assistance Anwendungen ermöglicht, wie wir es von Cloud Plattformen kennen, auf einer verteilten, stadtweiten Edge Infrastruktur. CognitiveAR ist unser erstes Projekt, und ist die Basis für eine Komplettlösung für AI-gestützte Augmented Reality (AR) Anwendungen mit Echtzeitanforderungen. AR Entwickler\_Innen können mit unserer Plattform eine Vielzahl von assistiven Anwendungen für tragbare AR Geräte wie etwa AR Brillen umsetzen. Unsere Plattform und Softwarebibliotheken vereinfachen zwei Kernfunktionalitäten solcher Anwendungen: das

Tracking von Personen oder Objekten in der Umgebung, sowie AI-basierte Analyse von Kamera-feeds der AR Geräte auf neuartiger Edge Computing Infrastruktur. Dieses Projekt baut auf mehreren Jahren an Forschungsergebnissen und -prototypen auf und demonstriert wie AR Entwickler\_innen ein transparenter Zugriff auf eine edge computing Plattform ermöglicht werden kann.

## 3 Verlauf der Arbeitspakete

### 3.1 **Arbeitspaket 1 - AP1: Detailplanung und Formales am Projektstart**

Der Projektstart wurde rasch und ohne nennenswerte Probleme abgewickelt.

### 3.2 **Arbeitspaket 2 - AP2: Design & Technologie Evaluation**

Unsere initiale Fragestellungen waren: a) gibt es existierenden KI-basierten Videoanalyse Technologien und wenn ja wie sind diese nutzbar? b) braucht man wirklich Edge Computing, oder könnte man unsere Anwendungsfälle auch mit Cloud-Basierten Ansätzen lösen? c) Welche Netzwerkprotokolle eignen sich für KI-basierte Videoanalyse? d) Welche Lösungen gibt es zum tracken von Objekten im 3D-Raum anhand von 2D Kamerabildern?

Die Antworten auf diese Fragen konnten wir durch eine wissenschaftliche und systematische Herangehensweise gut erarbeiten, und haben diese sogar als Paper auf der IEEE VR Publiziert (siehe Öffentlichkeitsarbeit).

### 3.3 **Arbeitspaket 3 - AP3: Computer Vision / Artificial Intelligence Komponenten**

Viele Anwendungsfälle in unserer Problemdomäne erfordern das Lokalisieren von Objekten im 3D Raum anhand von statischen Kameras. Etwa zum Positionieren von Autos um deren Positionen an AR Geräte in der Nähe zu senden. Diesen wesentlichen Teil von AP3 haben wir durch unser System Cyber-Physical Object Positioning fertig. Wir verwenden dafür OpenCV und diverse KI Modelle für die Tiefenerkennung. Wir können Objekte in Echtzeit erkennen und die Positionsdaten von 2D in 3D Positionen übersetzen. Diese werden über das Netzwerk an umliegende Geräte gesendet. Ein Demonstration dazu haben wir in unserer "Proof of Concept Anwendung". Obwohl damit mit AP3 Ende im Plan waren, hat CPOP unter gewissen Annahmen funktioniert (z.B. dass Objekte am Boden stehen). Wir haben CPOP anschließend erweitert, dass es auch frei-fliegende Objekte im Raum tracken kann, wie etwa hier zu sehen: <https://www.youtube.com/watch?v=1p5PHNfvsHY>

### 3.4 **Arbeitspaket 4 - AP4: Plattform Implementierung**

Ein wesentlicher Bestandteil von AP3 und AP4 ist ein Server für KI-basierte Echtzeit- Videoanalyse zum verarbeiten von Kamera-Feeds (sowohl statische Kameras als auch der AR Geräte), der sich

auf kleinen Edge-Computing Geräten (wie etwa dem NVIDIA Jetson Xavier NX) betreiben lässt. Unsere Plattform CogStream implementiert eine Plugin- Architektur, mit der sogenannte Engines verwaltet werden können. Engines kapseln die Funktionalität eines oder mehrere KI Modelle um die Bilder der Kamera zu analysieren. Sowohl das Framework, der Server, als auch die Engines befinden sich im selben Repository, dabei sind Framework und Server dem AP4, und die Engines dem AP3 zuzuordnen. Wir haben bereits drei Engines implementiert: Objekterkennung, Gesichtserkennung, und Facial Expression Recognition (Emotionserkennung). Die HoloLens kann sich bereits zum Server verbinden um den Kamera-Feed der Brille analysieren zu lassen.

### **3.5 Arbeitspaket 5 - AP5: Augmented Reality Apps**

Als Projektergebnis 9 haben wir ein Client SDK definiert, mit dem es AR Entwickler\_Innen leicht fällt mit unserem System zu kommunizieren. Diese Entwicklung hat im wesentlichen in AP5 stattgefunden. Der C# Code der es einer Augmented Reality App erlaubt sich mit der Plattform zu verbinden und mit den von der Plattform zur Verfügung gestellten Engines zu kommunizieren befindet sich in folgendem Repository: <https://github.com/cognitivexr/cognitivexr-unity-client> . Der Code kann von anderen AR Entwickler\_Innen heruntergeladen und als App-Package in ein Unity-Projekt importiert werden. Diese Herangehensweise hat sich für uns als produktiv herausgestellt, führte aber auch dazu, dass wir noch nicht in der Lage waren mit unserem Client SDK zu werben, und daher ist unsere open source traction auch noch sehr niedrig.

Mittels CPOP haben wir bereits eine simple Demo Anwendung aus AP5 umgesetzt: Augmented Reality Objekt-Tracking mittels Cyber-Physical Object Positioning (CPop). Video: <https://www.youtube.com/watch?v=nY3PLUTVSbw>

Die Client-seitige AR Demo App ist als Unity-Applikation in folgendem Repository verfügbar: <https://github.com/cognitivexr/unity-demo-app/>

Vollständig implementiert ist auch eine weitere Anwendung die das Video offloading mit AP4 demonstriert. Die HoloLens sendet dabei ihren Kamera-Feed zu einem Edge Knoten, der Gesichter und Emotionen erkennt, und im Display die Information entsprechend Anzeigt. Eine Demo Video der Anwendung ist hier verfügbar: <https://github.com/cognitivexr/cognitivear-demo#app-2-video-offloading-for-facial-expression-recognition>

Ein Blogartikel zum Client SDK und den Demo-Applikationen wurde ebenfalls veröffentlicht.

Eine Herausforderung war das Anschaffen zusätzlicher AR Brillen neben der HoloLens. Viele Anbieter haben Lieferschwierigkeiten bzw sind viele Brillen noch in Entwicklung und gar nicht Verfügbar, obwohl das im Marketing oft suggeriert wird. Unser Projekt erzielt daher den minimalen Erfolg (HoloLens Unterstützung).

### **3.6 Arbeitspaket 6 - AP6: Demo-Deployment**

Um Projektergebnis 8 (Infrastruktur Prototyp) und 11 (End-To-End Demo) zu erreichen, haben wir in den Räumlichkeiten der TU Wien ein kleines Labor einrichten können. Dort haben wir diverse

Hardware die wir von der TU Wien nutzen können aufgebaut (etwa eine HoloLens, und Hardware für Edge Knoten). Für die Bauanleitung des Edge Knotens haben wir bereits zwei Blogeinträge geschrieben (Siehe Publikationen und PR), und ein Repository angelegt das sowohl die Anleitung als auch Scripts zum Automatisieren der Konfiguration enthält.

Outdoor Umsetzung (z.B. erkennen von Autos, oder Montierung von Kameras auf Ampeln) hat sich im Laufe der Entwicklung durch technische Herausforderungen als nicht im Projektumfang möglich herausgestellt. Hier erreichen wir daher den minimal Erfolg mit einem Indoor Prototypen.

### 3.7 Arbeitspaket 7 - AP7: Dissemination

Siehe Öffentlichkeitsarbeit/Vernetzung.

### 3.8 Arbeitspaket 8 - AP8: Dokumentation und Formales am Projektende

Mit diesem AP wurden die formalen Kriterien des Projektendes erfüllt. Sowohl Endbericht als auch Userdokumentation wurden in kurzer Zeit umgesetzt. Darüber hinaus hat uns das Zusammenfassen der Projektergebnisse dabei geholfen über den Projektausgang und -erfolg reflektieren, den wir als überaus positiv empfinden.

## 4 Umsetzung Förderauflagen

Die Förderauflagen zu unserem Projekt wurden bei Projektanfang mit einer Ergänzung zum Projektantrag welche Inhaltliche und Personelle Unklarheiten aufgeklärt hat, und am 23.10.2020 an Dr. Andreas Koman gesendet wurde erfüllt.

## 5 Liste Projektergebnisse

1	Projektzwischenbericht	CC-BY-SA-4.0	<a href="https://www.netidee.at/cognitivear/">https://www.netidee.at/cognitivear/</a>
2	Projektendbericht	CC-BY-SA-4.0	<a href="https://www.netidee.at/cognitivear/">https://www.netidee.at/cognitivear/</a>
3	Entwickler_innen-DOKUMENTATION	CC-BY-SA-4.0	<a href="https://www.netidee.at/cognitivear/">https://www.netidee.at/cognitivear/</a>
4	Anwender_innen-DOKUMENTATION	CC-BY-SA-4.0	<a href="https://www.netidee.at/cognitivear/">https://www.netidee.at/cognitivear/</a>
5	Veröffentlichungsfähiger Einseiter	CC-BY-SA-4.0	<a href="https://www.netidee.at/cognitivear/">https://www.netidee.at/cognitivear/</a>

6	Externkommunikation	CC-BY-SA-4.0	<a href="https://www.netidee.at/cognitivear/">https://www.netidee.at/cognitivear/</a>
7	Die <b>Laufzeitplattform</b> besteht aus zwei wesentlichen Komponenten: <b>CogStream</b> für die AI-basierte Echtzeitanalyse von Kamerafeeds von AR Geräten, und dem <b>CPOP Server</b> , der Objekt-positionsdaten aus der Umgebung an AR Geräte sendet.	Apache 2.0	<a href="https://github.com/cognitivexr/cognitivexr-platform">https://github.com/cognitivexr/cognitivexr-platform</a>
8	Der <b>Infrastruktur Prototyp</b> ist eine Bau- und Installationsanleitung für eine „ <b>Edge Node</b> “ die aus einer Kamera und mehreren Rechnerknoten besteht um die Laufzeitplattform zu betreiben.	CC-BY-SA-4.0	<a href="https://github.com/cognitivexr/edge-node">https://github.com/cognitivexr/edge-node</a>
9	Die <b>Software Clients</b> werden von EntwicklerInnen verwendet um sich mit der Plattform zu verbinden und AR Apps zu entwickeln.	Apache 2.0	<a href="https://github.com/cognitivexr/cognitivexr-clients">https://github.com/cognitivexr/cognitivexr-clients</a>
10	Unsere <b>AR App</b> dient als Beispiel, wie man mit unserer Plattform eine Unity-basierte HoloLens 2 Anwendung für Personen Tracking und Emotionserkennung entwickeln kann.	GPL 3.0	<a href="https://github.com/cognitivexr/unity-demo-app">https://github.com/cognitivexr/unity-demo-app</a>
11	Unsere <b>End-to-end Demo</b> zeigt die komplette Integration der Ergebnisse 7-10, und besteht aus Videos, einer Anleitung, so wie Source Code um die Komponenten zu starten.	Apache 2.0	<a href="https://github.com/cognitivexr/cognitivear-demo">https://github.com/cognitivexr/cognitivear-demo</a>

## 6 Verwertung der Projektergebnisse in der Praxis

Sowohl unsere Plattform Software als auch die Demo App befinden sich im Prototyp stadium. Jedoch kann man mit der umfangreichen Anleitung zum Starten der Plattform, und der verfügbaren Client Software bereits einiges erreichen.

Hierbei gibt es mehrere Zielgruppen:

AR Entwickler\_Innen können mit CognitiveAR und mit unserer Demo App als Vorlage bereits eigene Prototypen mit Unity entwickeln. Eine App die wir Beispielsweise angedacht haben dient zum Erkennen von Zeichensprache, wo eine AR Brille die Handzeichen in eingeblendeten Text übersetzt. Da unsere Client Software für Unity konzipiert ist, und viele AR Geräte Unity unterstützen, gibt es hier eine niedrige Barriere für den Einstieg.

Forscher\_Innen im Bereich Edge Computing können CogStream als Anwendungsfall nutzen. Durch die Neuartigkeit des Bereichs wird in Forschungsarbeiten selten mit greifbare Anwen-

dungsfällen evaluiert. CogStream bietet hier mit dem offloaden von Videoanalyse Tasks von Edge Geräten auf GPU-beschleunigte Infrastruktur ein interessantes Umfeld für Evaluierungen. CogStream wird an der Distributed Systems Group der TU Wien bereits zur Evaluierung von Algorithmen für die Systemorchestrierung verwendet.

CPOP ließe sich auch gesondert verwenden um beispielsweise die Distanz von Personen mit CCTV Kameras zu Analysieren. Ein denkbarer Einsatzort wären hier etwa der Einstieg zu Bergbahnen, um sicher zu stellen, dass Personen den durch COVID notwendigen Sicherheitsabstand einhalten.

## 7 Öffentlichkeitsarbeit/ Vernetzung

Die folgenden Publikationen, Blog Artikel, und PR Materialien wurden im Rahmen des Projekts veröffentlicht:

- [Radio Interview zu Cognitive AR auf “Start Me Up” Radio Magazin](#)
- **Fünf** Inhaltliche Blogartikel auf der **Netidee Community** Website:
  - [Announcing CognitiveXR](#)
  - [Systemdesign, erste Prototypen und Ergebnisse](#)
  - [Unsere Fortschritte im CognitiveAR Labor](#)
  - [Emotionserkennung auf der HoloLens mit CognitiveXR](#)
  - [CognitiveAR Projektabschluss](#)
- **Zwei** Blogartikel mit Technischen Inhalten auf der CognitiveXR Website
  - [Computer Vision with PyTorch on the Jetson Xavier NX](#)
  - [Setting up an Intel RealSense Camera on an NVIDIA Jetson Xavier NX](#)
- [Ein Vortrag und Paper zu CognitiveAR auf dem IEEE VR DISCE 2021 Workshop](#)
  - Dieser Vortrag hat zu einem Meeting für eine potentielle Kollaboration mit der UCLA und dem [OpenPTrack Team](#) geführt
- Social Media:
  - 172 follower auf [LinkedIn](#)
  - 39 follower auf [Twitter](#)
  - [YouTube channel](#) mit insgesamt über 600 Video views und 14 subscribers

Des weiteren planen wir in Kooperation mit dem VRVis zwei weitere Publikationen: eine zur AR App, und eine zum Mechanismus der Echtzeit-Objekterkennung von CPOP.

## 8 Eigene Projektwebsite

Projektwebsite: <https://cognitivxr.at>

Code Repository auf GitHub: <https://github.com/cognitivxr>

## 9 Geplante Aktivitäten nach netidee-Projektende

Zunächst planen wir die Verwertung der Resultate in eine weitere wissenschaftliche Publikation. Mit dem Feedback von AR Expert\_Innen möchten wir die Konzeptionierung neuer Features vorantreiben. Außerdem werden wir vorhandene Ideen zu kommerzialisierbaren Anwendungsfällen weiterentwickeln.

## 10 Anregungen für Weiterentwicklungen durch Dritte

Insbesondere durch die aktuellen Entwicklungen im Extended Reality (XR) Bereich werden Lösungen wie unsere immer relevanter. Es lassen sich mit unserer Plattform eine Vielzahl an AR Anwendungen umsetzen, etwa ein Assistent für das Übersetzen von Zeichensprache, oder Hybrid-Reality (wie etwa Microsoft Mesh). Die Integration mit anderen Geräten, etwa die Nreal Light wäre eine spannende Weiterentwicklungsmöglichkeit.