Evaluationsbericht:

Safer Internet-Bot

Wien, November 2018

Österreichisches Institut für angewandte Telekommunikation

Impressum

Mag.ª Louise Horvath

Österreichisches Institut für angewandte Telekommunikation

Ungargasse 64-66/3/404

1030 Wien

Inhalt

Einleitung

Kinder nehmen Beratungsangebote über Telefon oder E-Mail immer weniger wahr. Tritt in ihrem Online-Alltag ein Problem auf, wollen sie in möglichst wenigen Klicks die passende Lösung finden. Um diesen Anforderungen gerecht zu werden, können Chatbots eine technische Option sein. Statt personalintensiver Lösungen, beantworten textbasierte Dialogsysteme wiederkehrende Fragen zeitnah.

Wie viel taugen Chatbots tatsächlich, wenn es um die Beratung von Kindern geht?

Im Forschungs- und Entwicklungsprojekt „Safer Internet-Bot" wurde ein Chatbot für Kinder im Volksschulalter entwickelt. Das Ziel war, dass er eingesandte Kettenbriefe erkennt und Fragen dazu beantwortet. Im Dezember 2017 begann der Entwicklungsprozess – seit Oktober 2018 läuft der Chatbot.

Damit Anfragen an den Chatbot möglichst richtig zugeordnet werden können, braucht es eine gute Datenbasis. Das seit 2016 für 1,5 Jahre betriebene Service „Kettenbrief-Telefon" wurde als Grundlage für das Bauen des Dialogsystems genutzt. In der Woche waren 400-500 Chats von Kindern gestartet worden – das Saferinternet.at-Team hatte zunehmend standardisiert geantwortet. Dieses Archiv an Dialogen wurde anonymisiert ausgelesen und kategorisiert.

Dieser Bericht evaluiert den Entwicklungsprozess des Safer Internet-Bots. Der Chatbot beruht im Wesentlichen auf zwei Komponenten: Die Anbindung an Messenger Plattformen (Kapitel 1), sowie das Dialogsystem (Kapitel 2). Es folgt eine Evaluation des laufenden Betriebs des Safer Internet-Bots (Kapitel 3).

Darauf aufbauend werden die zentralen Lektionen für den Transfer der Erfahrungen in andere Bereiche zusammengefasst (Kapitel 4).

Kettenbriefe sind für Kinder im Volksschulalter herausfordernd: täglich erhalten sie über soziale Medien diese Nachrichten, die ihnen Angst oder falsche Hoffnungen machen sollen.

Im Jahr 2016 richtete Saferinternet.at deshalb das „Kettenbrief-Telefon" ein. Kinder konnten ihre Kettenbriefe an das Team weiterleiten und erhielten eine kurze Rückmeldung.

1. Die Anbindung

Nicht jeder Messenger-Dienst eignet sich gleichermaßen für den Einsatz von einem Chatbot. Die Frage der Anbindung stellte sich deshalb als größte Herausforderung, denn Kinder im Volksschulalter nutzen derzeit vor allem WhatsApp. Dieser Dienst lässt bislang aber keine Anbindung von Chatbots zu.

Im Safer Internet-Bot Projekt experimentierte das Team deshalb auch mit einer Anbindung an WhatsApp, bei der über einen virtuellen Server auf die Web-Version zugegriffen wird. Da sich Elemente in WhatsApp regelmäßig verändern, arbeitet das System mit einer händischen Kalibrierung der Elemente und simuliert das langsame Antworten[1]. Diese Anbindung wird derzeit nur intern genutzt, denn sie ist noch zu fehleranfällig. Die **reguläre Anbindung des Safer Internet-Bots** läuft über den Messenger-Dienst **Telegram**. Eine Erweiterung der Anbindungen ist möglich – angedacht ist zum Beispiel den Chatbot für den Facebook Messenger freischalten zu lassen.

2. Das Chatbot-Framework und LUIS

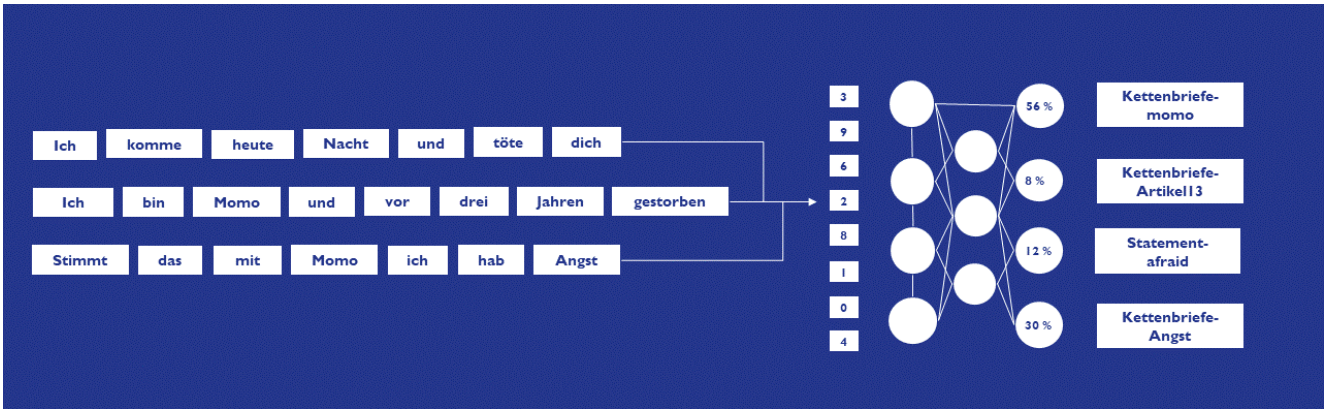

Abbildung 1: Natural Language Processing erklärt anhand von Anfragen an den Safer Internet-Bot. Bild: CC-BY (Saferinternet.at)

Um Anfragen zu bearbeiten wird mit dem LUIS System von Microsoft gearbeitet. Dabei werden eingehende Anfragen durch Natural Language Processing zerlegt

[1] https://github.com/robert4os/wac

und analysiert. Errechnet wird vom System mit welcher Wahrscheinlichkeit eine Anfrage einem bestimmten „Intent“ zugewiesen werden kann.

3. Das Dialogsystem

Das System umfasst derzeit **43 „Intents“** – sie sollen ermöglichen jede Anfrage an den Chatbot passend einzuordnen. Intents (engl. für Absicht) sind Aktionen, die ein/e Nutzer/in ausführen möchte. Jedem Intent ist zumindest eine „Utterance“ zugeordnet. Für den Intent „question-saferinternet“, wäre das z.B. die Frage „Was ist Safer Internet?“. Die Intents werden abhängig von den eingehenden Anfragen laufend ausgebaut und verfeinert.

Das System des Chatbot unterscheidet zwischen allgemeinen Anfragekategorien („Intents“), den konkreten Anfragen („Utterances“) und den für jede Kategorie erstellten Antworten.

In Abbildung 2 ist ein klassischer Prozess beim Safer Internet-Bot dargestellt. Ein Kind stellt eine Anfrage zum Hoax Momo – erkannt wird, dass damit der Intent „chainletter-momo“ geäußert wird und entsprechend wir eine passende Antwort zugeordnet.

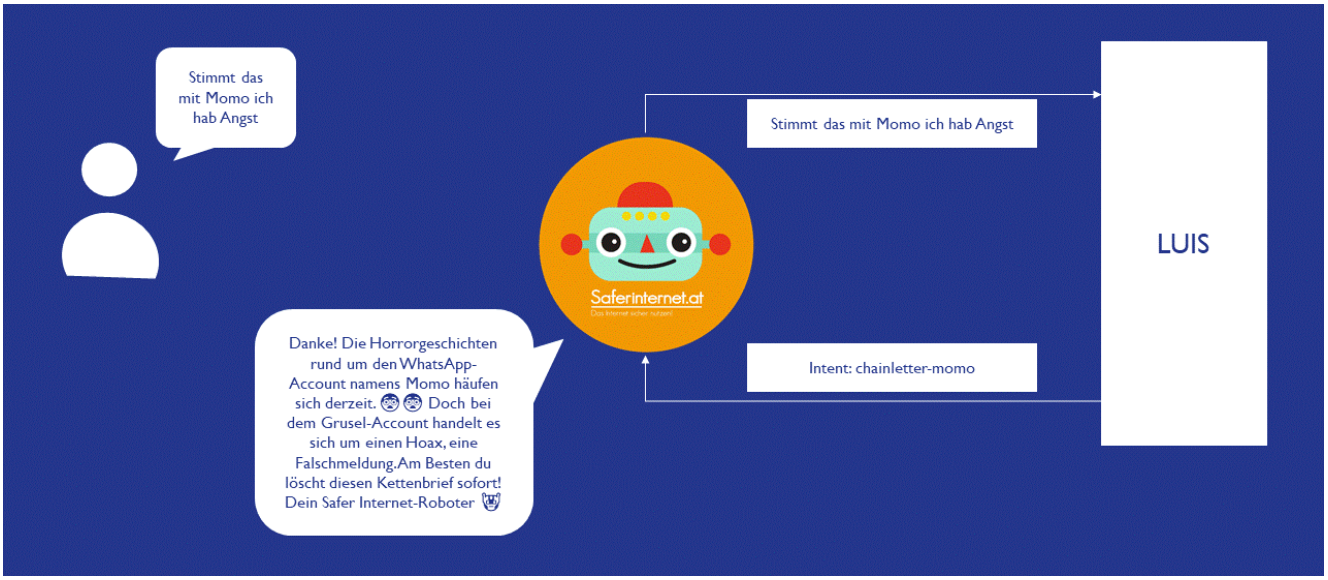

Abbildung 2: Der Ablauf von der Anfrage hin zur Antwort beim Safer Internet-Bot. Bild: CC-BY (Saferinternet.at)

Im Folgenden werden die Intents des Safer Internet-Bots beschrieben und Beispiele für zugeordnete Utterances gegeben.

3.1. Die Kettenbriefe

Saferinternet.at hatte bereits im Zuge des „**Kettenbrief-Telefon**“ im Jahr 2016 begonnen die erhaltenen Kettenbriefe in Kategorien einzuteilen. Die Absicht dabei war unterscheiden zu können, welche harmlos und welche angsteinflößend sind und möglichst passende, standardisierte Antworten zu geben.

Die Intents des Chatbots sind noch umfassender geworden und illustrieren auf welche Datenbasis das Dialogsystem fußt. Insgesamt sind **617 Kettenbriefe** im Archiv. Das Gros sind die allgemeinen (harmlosen) Kettenbriefe, in diesen wird nichts Besonderes versprochen – weder Belohnung, noch Sanktion.

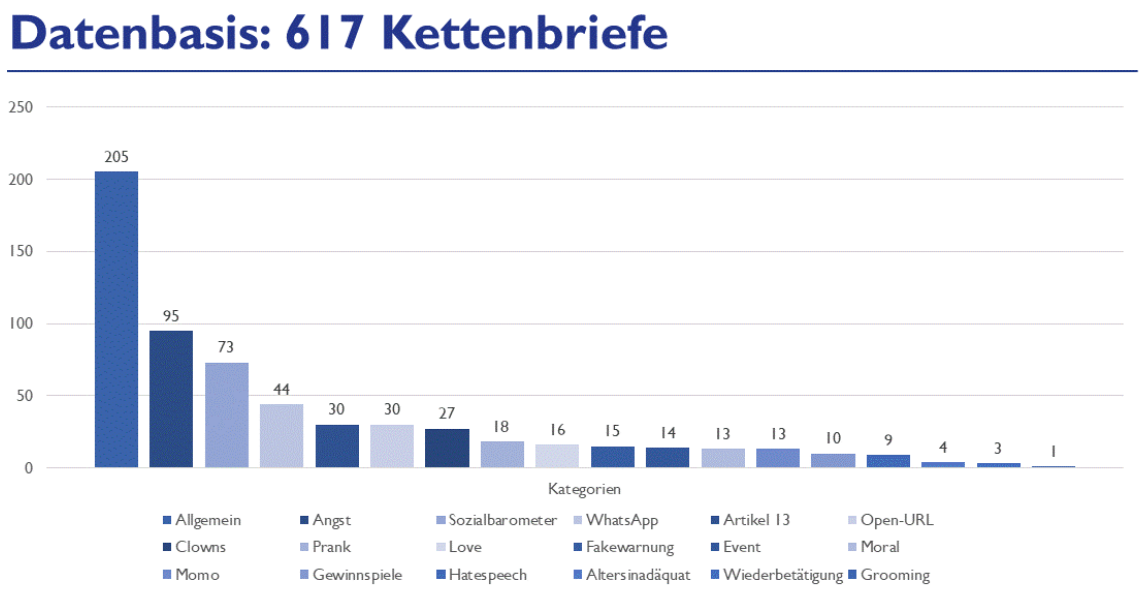

Abbildung 3: Kettenbriefe im LUIS-System, Stand: Mitte November 2018.
Bild: CC-BY (Saferinternet.at)

Manche der Intents wie z.B. zu Kettenbriefen mit nationalsozialistischer Wiederbetätigung oder Grooming gibt es wenige Beispiele (Utterances), aber sie sind mindestens ebenso bedeutend. **Utterances werden nur einem Intent zugeordnet**. Für einzelne Kettenbriefe, die besondere Fragen ausgelöst haben wie z.B. „Momo“ wurde eine eigene Kategorie geschaffen, damit die Antworten auch möglichst passend gegeben werden können.

Intents	Labeled Utterances
Chainletter-ageunsuitable	4
Chainletter-clown	25
Chainletter-event	13
Chainletter-fakewarnung	13
Chainletter-general	193
Chainletter-grooming	1
Chainletter-hatespeech	9
Chainletter-lottery	10
Chainletter-momo	13
Chainletter-moral	12
Chainletter-prank	16
Chainletter-scary	87
Chainletter-socialbarometer	73
Chainletter-whatsapp	44
Chainletter-wiederbetaetigung	3
Statement-openurl	30

Je nachdem wie spezifisch die Antwort auf einen Kettenbrief ausfallen soll, werden Intents formuliert. Im Folgenden als Beispiele die Kettenbriefe zu Artikel 13.

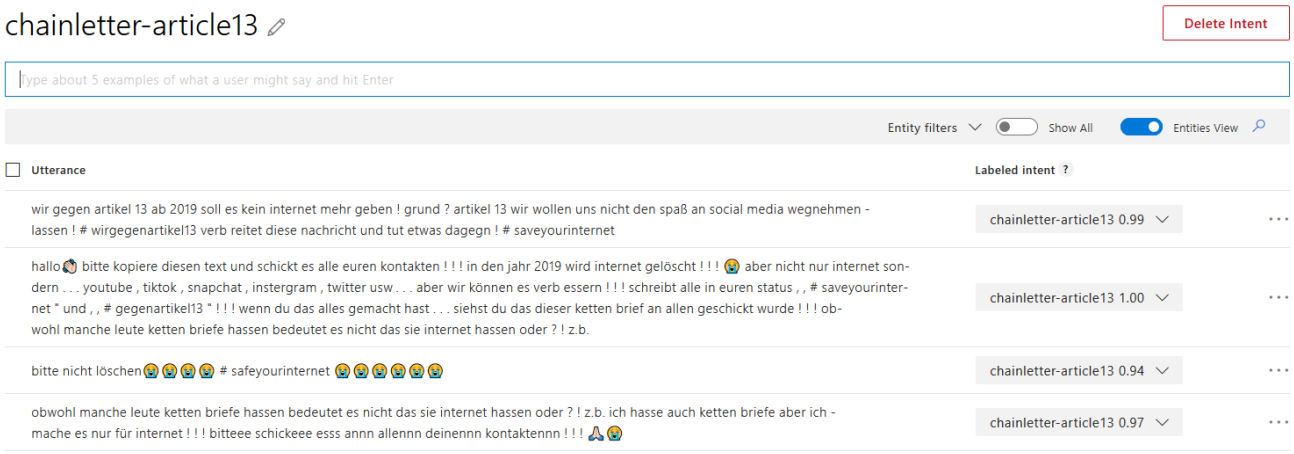

Abbildung 4: "Artikel 13" Intent. Bild: CC-BY (Saferinternet.at)

Dem Intent „chainletter-article13“ sind alle Kettenbriefe als Utterances zugeordnet. Rechts zu sehen ist mit welcher Wahrscheinlichkeit das System diese dem Intent zuweist (max. 1).

3.2. Die Fragen

Wenngleich der Fokus auf der Beratung zu Kettenbriefen liegt, werden etliche weitere Fragen beantwortet.

Dabei geht es zum einen um die **Vorstellung des Chatbot** selbst, im Sinne einer transparenten Darstellung davon, wer die Anfragen beantwortet („Wer bist du?“, „Was machst du?“, „Wie geht’s dir?“) und kindgerechter Erklärungen zu seinen Funktionen („Was ist ein Chatbot?“).

Dazu kommen **Fragen zu Kettenbriefen** und nach deren Wahrheitsgehalt. Konkret geht es um die Fälle, in denen Kinder zuerst einen Kettenbrief schicken und als nächste Nachricht fragen, ob dieser stimmt. Das Problem dabei ist, dass der Safer Internet-Bot in diesem Fall zuerst auf den Kettenbrief antwortet und erst als nächstes auf die Nachricht „Stimmt das?“ – allerdings ohne diese Anfrage mit dem eingeschickten Kettenbrief verbinden zu können. Es wurde deshalb eine Kategorie erstellt, die es erlauben soll Kindern an dieser Stelle zu antworten und gleichzeitig klarzustellen, dass der Chatbot auf Kettenbriefe direkt antwortet („question-stimmtdas“).

Immer wieder gibt es auch Fragen dazu, warum es Kettenbriefe gibt, wer hinter Kettenbriefen steht und was eigentlich Kettenbriefe sind („question-wasisteinkettenbrief“).

Intents	Labeled Utterances
Question-bot	4
Question-bot2	4
Question-chainletter	12
Question-lottery	3
Question-organization	80
Question-saferinternet	8
Question-stimmtdas	33
Question-wasisteinkettenbrief	8

3.3. Der Gesprächsrahmen

Der Chatbot beantwortet jede eingelangte Frage einzeln – schreibt ein Kind zum Beispiel drei Nachrichten mit Begrüßungen, folgen darauf auch drei Antworten mit Begrüßungen des Chatbots. Kinder sollen allerdings nicht zu langen Gesprächen zu anderen Themen als Kettenbriefen verleitet werden. Die nicht-themenspezifischen Anfragen, die als Intents im Dialogsystem festgehalten werden, sind deshalb stark limitiert worden.

Utterances zum Intent „express-thanks“
den mag ich nicht
die hab ich auch noch
hier ein ketten brief den ich bekommen habe :
ich habe so ne ketten brief bekommen
ich schick euch das
ich sende euch die ketten briefe

In einigen Fällen wurde über Ausstiegsmöglichkeiten aus dem Dialog nachgedacht. Das mündete zum Beispiel in **Bedanken als Ausstieg aus Gesprächen** (Intent: „express-thanks“). Nach Analyse der Dialoge mit den Kindern wurde klar, dass viele Aussagen keine spezifischen Antworten mehr brauchen und die Gespräche an diesen Stellen mit einem Danke abgebrochen werden können. Unter dem Intent „thanks“ wird hingegen auf das Danke von Kindern geantwortet.

Kinder hatten sich zuvor schon beim Saferinternet.at-Team trotz generisch als Team signierter Antworten oft nach dem Befinden des Teams erkundigt oder auch mit **Liebeserklärungen** (z.B. „hab dich lieb“) reagiert. Diese wurden als eigene Intents aufgenommen und werden humorvoll beantwortet. Ähnlich das Prinzip bei **Beschimpfungen** aller Art.

Über den Intent „statement-banal“ soll ein **sanfter Ausstieg aus Gesprächen** ermöglicht werden, wenn Anfragen über das angebotene Service hinausgehen. Als Antwort wird ein lächelndes Emoji zurück gespielt – aus der Erfahrung mit dem „Kettenbrief-Telefon“ hatte sich dies als die einfachste, akzeptierte Form des Gesprächsabschluss herausgestellt.

Intents	Labeled Utterances
Bye	18
Express-thanks	17
Greeting	55
Love declarations	13
Statement-insult	41
Question-wiegehts	21
Statement-banal	71
thanks	21
Question-dasnervt	10
Statement-dumachstfehler	5

3.4. Wenn mehr Beratung gebraucht wird

Der Safer Internet-Bot kann keine Beratung leisten, die über die Beantwortung von Fragen zu Kettenbriefen hinausgeht bzw. psychologische Dimensionen hat. Wenn mehr Ratschläge gebraucht werden oder spürbar weitere Begleitung notwendig ist, verweist der Chatbot deshalb systematisch auf die psychologische Beratungsstelle „147 Rat auf Draht“.

Intents	Labeled Utterances
Question-advicenecessary	60
Statement-afraid	32
Question-wasistrataufdraht	13

3.5. Was nicht beantwortet werden kann

Der Intent „None“ ist für jeden Chatbot unerlässlich. Er wird genutzt um dem Chatbot beizubringen, welche Anfragen nicht für das Service als relevant zu erachten sind. Landet eine Anfrage bei None, kann der Chatbot z.B. weitere Fragen stellen oder nochmals seine Funktionen erklären.

Intents	Labeled Utterances
None	84

Der Safer Internet-Bot soll sich darauf beschränken zum Thema Kettenbriefe zu beraten – insofern fallen unter diese Kategorie von zufälligen Buchstabeneingaben hin zu anderen Anfragen wie z.B. „Welches Virenprogramm soll ich nutzen?“.

4. Der Betrieb des Safer Internet-Bots

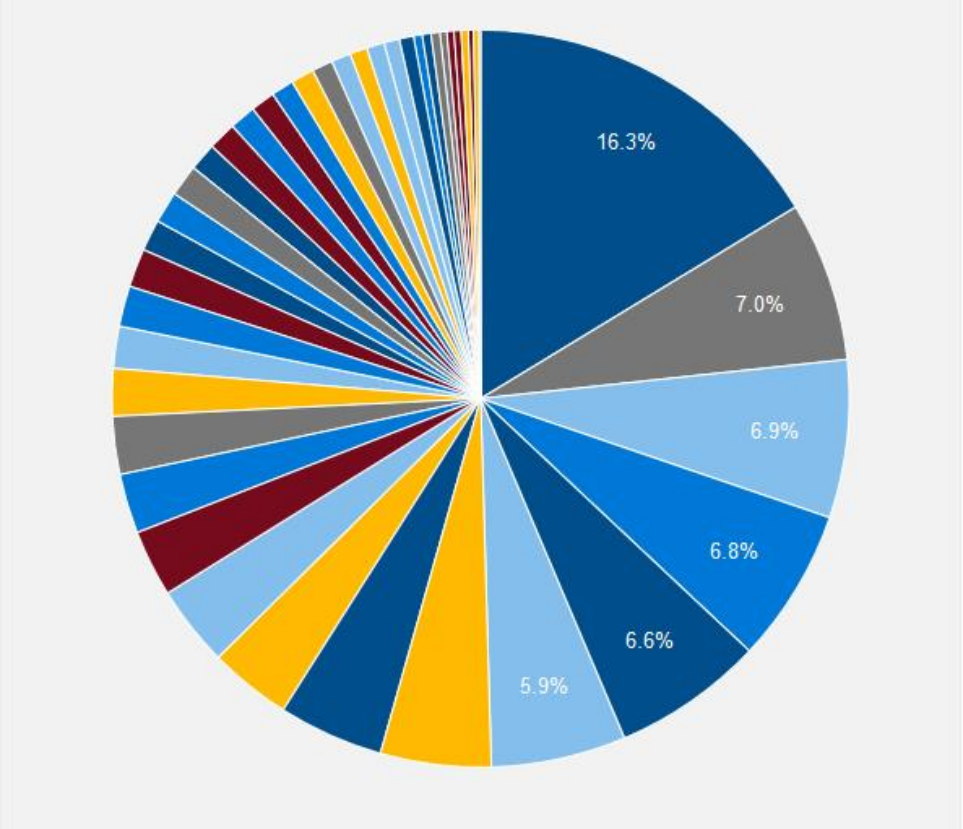

Abbildung 5: Mit jeder Anfrage verändert sich dieses Diagramm, in welchem angezeigt wird, welches Gewicht (prozentuell) welcher Intent einnimmt. Bild: CC-BY (Saferinternet.at)

Erhält der Chatbot einen Kettenbrief, wird dieser in LUIS eingelesen und in die entsprechende Kategorie (Intent) geteilt – das Projektteam erhält keine Angaben darüber, von welcher Nummer bzw. von welchem Chat welche Nachrichten eingeschickt werden. Gesehen wird, in welchem Ausmaß welche Intents abgefragt werden. In Abbildung 2 ist dies z.B. die Kategorie „chainletter-general“, gefolgt

von der Kategorie „chainletter-scary“. Dies wird es Saferinternet.at in Zukunft erlauben Trends bei den Anfragen zu erkennen und darauf rasch zu reagieren.

4.1. Monitoring-Möglichkeiten

Es gibt eine **Monitoring**-Möglichkeit beim Chatbot. Jene Botschaften, die neu sind bzw. einer weiteren Zusicherung brauchen, werden getrennt angezeigt. Das Projektteam kann ihre Zuteilung bestätigen oder verändern. An dieser Stelle kann der Lernprozess der Künstlichen Intelligenz vom Chatbot entscheidend beeinflusst werden. Es wird möglich die Erkennung zu verbessern, aber auch davon ausgehend z.B. zusätzlich notwendige Intents auszumachen.

Utterances	Aligned in
ist es ok wenn ich ein liebes ketten brief bekomme bz. suche das andere emoji oder so?	Question organization
wie kann ich es meiner Mutter erzählen? ich trau mich nicht	Question advice-necessary
ich habe das ketten brief nicht mehr. Ich werde versuchen es zu bekommen. lg elsa	Express-thanks

4.2. Dialoge

Abbildung 6: Werden Kettenbriefe eingeschickt, erkennt der Chatbot sie mit großer Treffsicherheit. Ebenfalls gut klappt das Erkennen von Signalen, die darauf hindeuten, dass ein Kind zusätzliche Beratung durch Menschen bräuchte. Bild: CC-BY (Saferinternet.at)

5. Schlussfolgerungen

Die wichtigsten Lernmomente im Entwicklungsprozess des Chatbots für die Beratung von Kindern im Volksschulalter betreffen die Vorbereitung (1), die Erstellung der Antwortkategorien (2) und ein klarer Blick auf die Grenzen und Möglichkeiten textbasierter Dialogsysteme (3).

5.1. Zentrale Erfolgsbedingungen

Die grundlegende Regelstruktur eines Chatbots zu erstellen, gelingt recht einfach – aber der Erfolg der textbasierten Dialogsysteme beruht auf dem vorhandenen **Wissen zu den Bedürfnissen der Nutzer/innen und wie sie diese ausdrücken** (Was wollen sie wissen? Wie stellen sie ihre Fragen?).

Diese Bedingungen waren für den Safer Internet-Bot sehr günstig. Mit dem Service „Kettenbrief-Telefon“, das im Jahr 2016 gestartet worden war, gab es schon eine **umfassende Datenbasis** auf welcher aufgebaut werden konnte. Das Service erlaubte es auch laufend zu testen, wie erstellte Antworten für den Chatbot in der Praxis ankommen bzw. welche Anfragen zu erwarten sein werden.

Die **Einbindung der Nutzer/innen** bei der Erstellung von Antworten ist unerlässlich. Im Zuge des Entwicklungsprozesses des Safer Internet-Bot fand in diesem Rahmen neben der laufenden Einbindung über das „Kettenbrief-Telefon“ ein Workshop mit Kindern der dritten Klasse Volksschule statt.

In diesem wurde mit ihnen über das Thema Kettenbriefe gesprochen und in Kleingruppen wurden sie gebeten ihrer Kreativität freien Lauf zu lassen und auf Kettenbriefe zu antworten. Ihnen wurden auch Antworten aus dem bisherigen „Kettenbrief-Telefon“ sowie aus den ersten, entwickelten Chatbot-Antworten vorgelegt und sie wurden um ihre Meinung dazu gebeten (Was verstehe ich nicht? Was finde ich gut? Was gefällt mir weniger?).

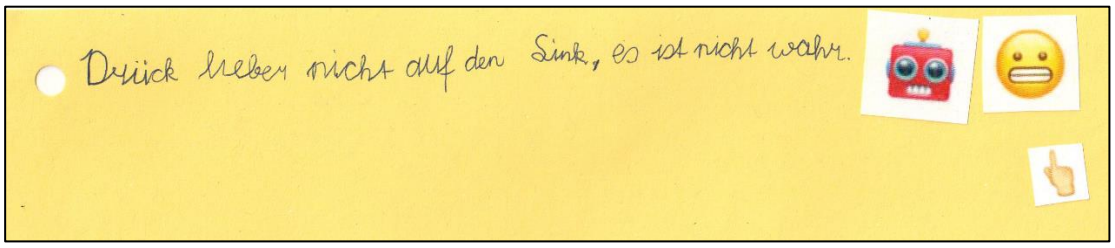

Abbildung 7: Chatbot-Antwort zu einem dubiosen Link in einem Kettenbrief. Bild: CC-BY (Saferinternet.at)

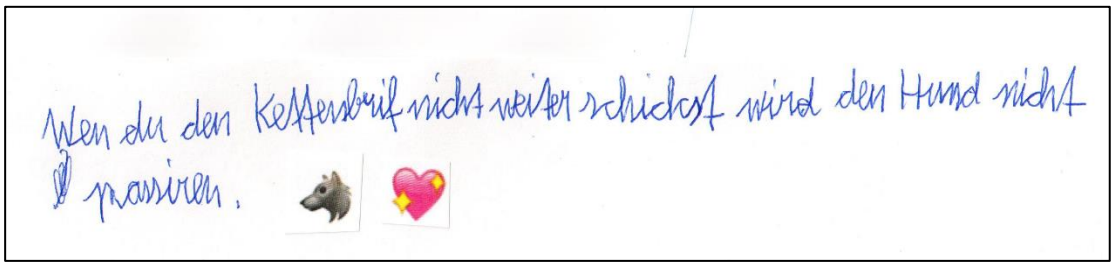

Abbildung 8: Chatbot-Antwort auf einen Kettenbrief mit einer Fake News zu Tierquälerei. Bild: CC-BY (Saferinternet.at)

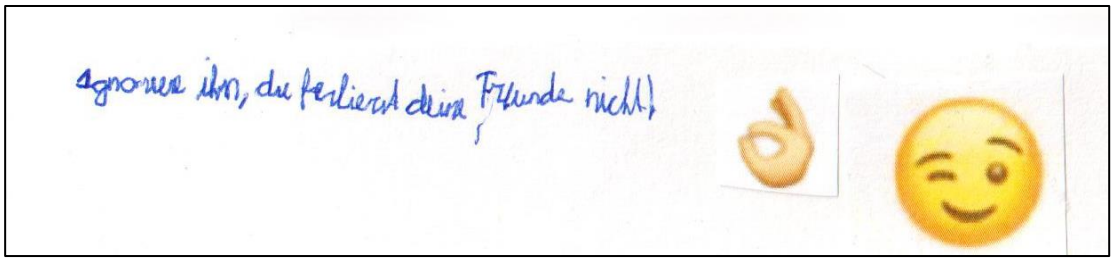

Abbildung 9: Antwort zu einem Kettenbrief, der damit droht, dass Freund/innen verloren werden könnten. Bild: CC-BY (Saferinternet.at)

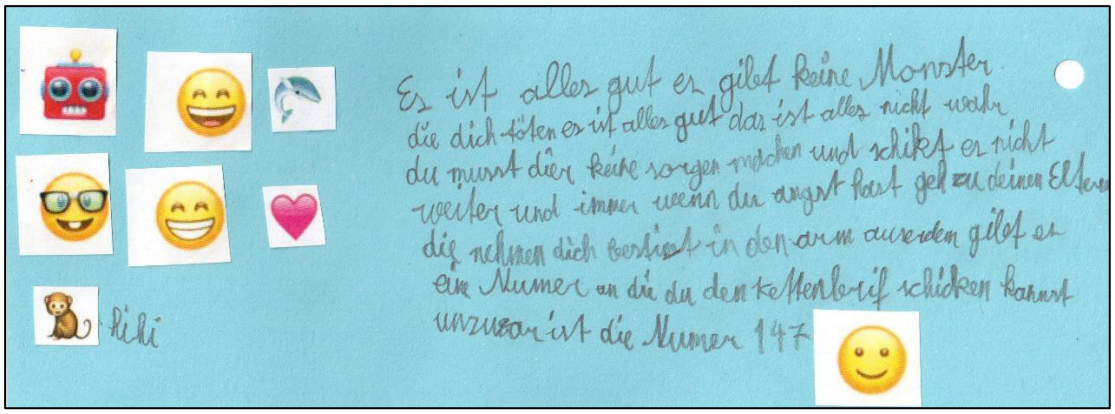

Abbildung 10: Antwort zu einem angsteinflößenden Kettenbrief. Bild: CC-BY (Saferinternet.at)

5.2. Ausblick

Der Safer Internet-Bot war in seiner Entwicklung für das ÖIAT eine Exploration der Chancen und Grenzen von semi-automatisierter Beratung auch in Hinblick auf andere Themen und Zielgruppen. Im Wissen um die sich wandelnden Anforderungen von digitalen Bürger/innen hinsichtlich von Beratung online, braucht es neue Formen und Zugänge der Vermittlung. Chatbots können, zeigt das Projekt, ein Weg seinen einen Teil der Beratungsleistung zu automatisieren.

Der Safer Internet-Bot zeigt dabei, wie wichtig es ist eine klare und eingegrenzte Fragestellung zu haben. Der Fokus auf die Beratung zu Kettenbriefen ermöglichte es einerseits früh mit einer Kategorienbildung zu beginnen, andererseits aber auch die Dialoge in ihrer Komplexität massiv zu reduzieren und die Treffsicherheit des Chatbots bei der Beratung zu erhöhen.

Der Safer Internet-Bot wird weiter als Service bei Safer Internet aufrechterhalten und ausgebaut werden.