Microsoft

DIGITAL SAFETY

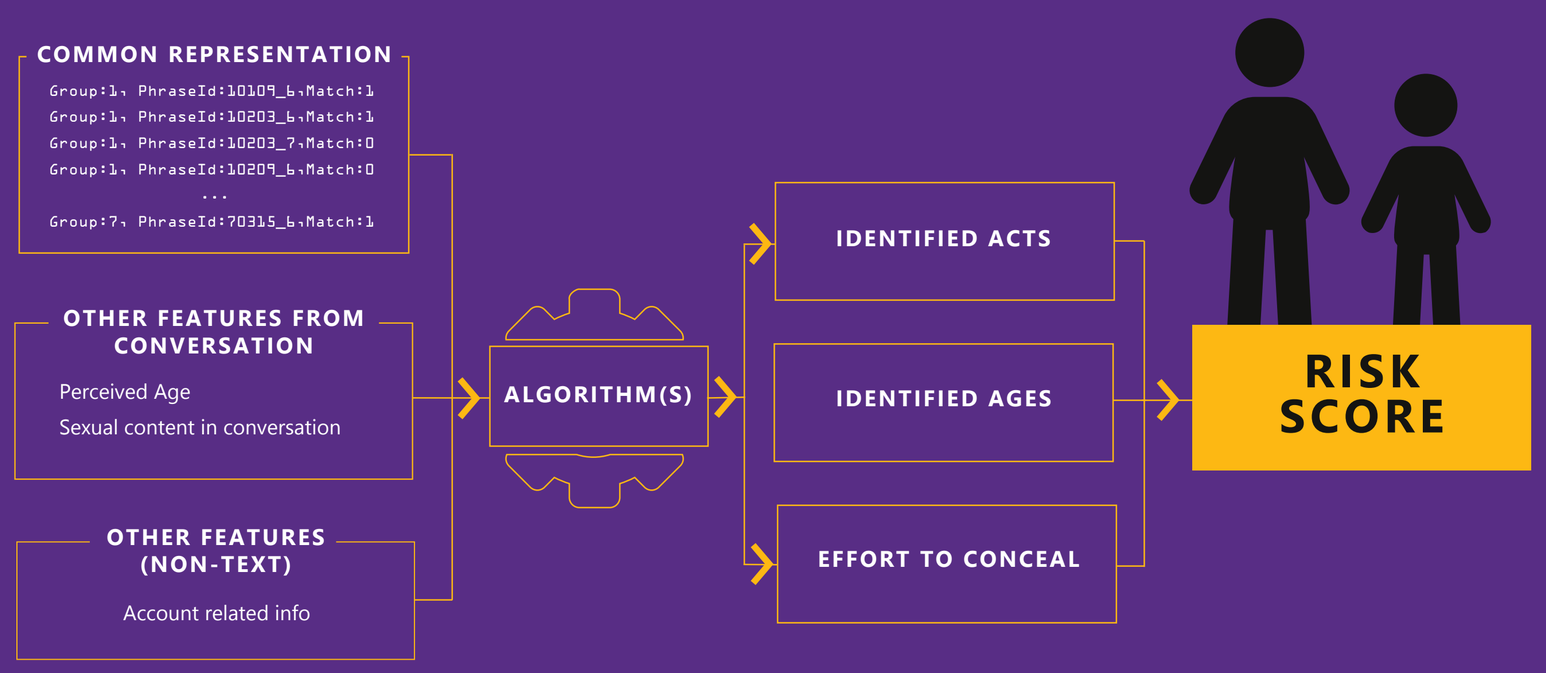
Detecting online child grooming for sexual purposes

Who?

Microsoft and the tech industry, specifically Kik, The Meet Group, Roblox and Xbox

What?

Creating and making freely available a new “technique” for detecting, addressing and reporting potential instances of child online grooming for sexual purposes



Where?

Online services offering chat can use the tool to identify potentially problematic conversations for later human review

When?

Since November 2018 when we held our “360” cross-industry hackathon at Microsoft HQ in Redmond, WA, USA

360 X-Industry Hack

Why?

To protect our customers and the integrity of our services by identifying and driving from our platforms predators who target children for sexual exploitation and abuse, and to report offenders to the National Center for Missing and Exploited Children

How it works

Building off a Microsoft patent, the technique is applied to historical text-based chat conversations

It evaluates and “rates” a series of characteristics in the conversations and assigns an overall probability rating

Rating can then be used as a determiner, set by individual companies, as to when flagged conversations should be sent to human moderators for review

Risky conversations filter

Not an issue

Unrelated act
Any 2 age groups

Any act
Adult/Adult

Relationship, Sexual Conversation, & Imagery Request/Exchange
Child/Child

Risky conversation, but in and of itself, not a major concern

In Person Meet
Child/Child

Imagery Request/ Exchange
Both Ages Inferred

High-risk conversation, but likely non-actionable

Sexual Conversation
Child pretending to be adult and adult **walks away** when known it is a child
One Age Known/One Inferred
Both Ages Inferred

Sexual Imagery Request
Child/Child: Not grooming, but other violation in most cases

Relationship
Child pretending to be adult and adult **walks away** when known it is a child
One Age Known/One Inferred
Both Ages Inferred

Sexual Imagery Request
Both Ages Inferred

Relationship
Both Ages Known

Imagery Request/ Exchange
Child pretending to be adult and adult **walks away** when known it is a child
One Age Known/One Inferred

Extreme risk; warrants actioning

In Person Meet
One Age Known/One Inferred
Both Ages Inferred

Sexual Imagery Exchange
Child pretending to be adult and adult **walks away** when known it is a child
One Age Known/One Inferred
Both Ages Inferred
Child/Child: Not grooming, but other violation in most cases

Sexual Imagery Request
Adult pretending to be child
Child pretending to be adult and adult **continues** when known it is a child
Child pretending to be adult and adult **walks away** when known it is a child
Both ages known
One Age Known/One Inferred

Imagery Request/Exchange
Adult pretending to be child
Child pretending to be adult and adult **continues** when known it is a child
Both Ages Known

Sexual Conversation
Adult pretending to be child
Child pretending to be adult and adult **continues** when known it is a child
Both Ages Known

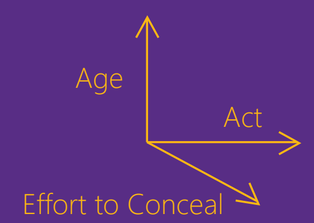
Relationship
Adult pretending to be child
Child pretending to be adult and adult **continues** when known it is a child

Imminent Risk

In Person Meet
Adult pretending to be child
Child pretending to be adult and adult **continues** when known it is a child
Both Ages Known

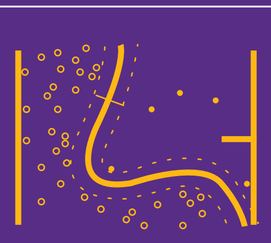
Sexual Imagery Exchange
Adult pretending to be child
Child pretending to be adult and adult **continues** when known it is a child
Both Ages Known

What we plan to deliver



1 A common set of threat dimensions to represent the risk of sexual grooming in chat

2 A common representation of conversations leveraging "regex" rules



3 Algorithms focused on identifying risk in a conversation and the possible ages of the participants

When we plan to deliver

Closed beta starting by end-2019. General free availability anticipated in first half of 2020 from Thorn **(www.wearethorn.org)**

THORN

Microsoft

Connect with us online!

Microsoft.com/SaferOnline
Facebook.com/SaferOnline
Twitter.com/Safer_Online
Youtube.com/MSFTOnlineSafety