# DAPter: Preventing User Data Abuse in Deep Learning Inference Services

Hao Wu[1], Xuejin Tian[1], Yuhang Gong[1], Xing Su[1], **Minghao Li[1 2]**, Fengyuan Xu[1*]

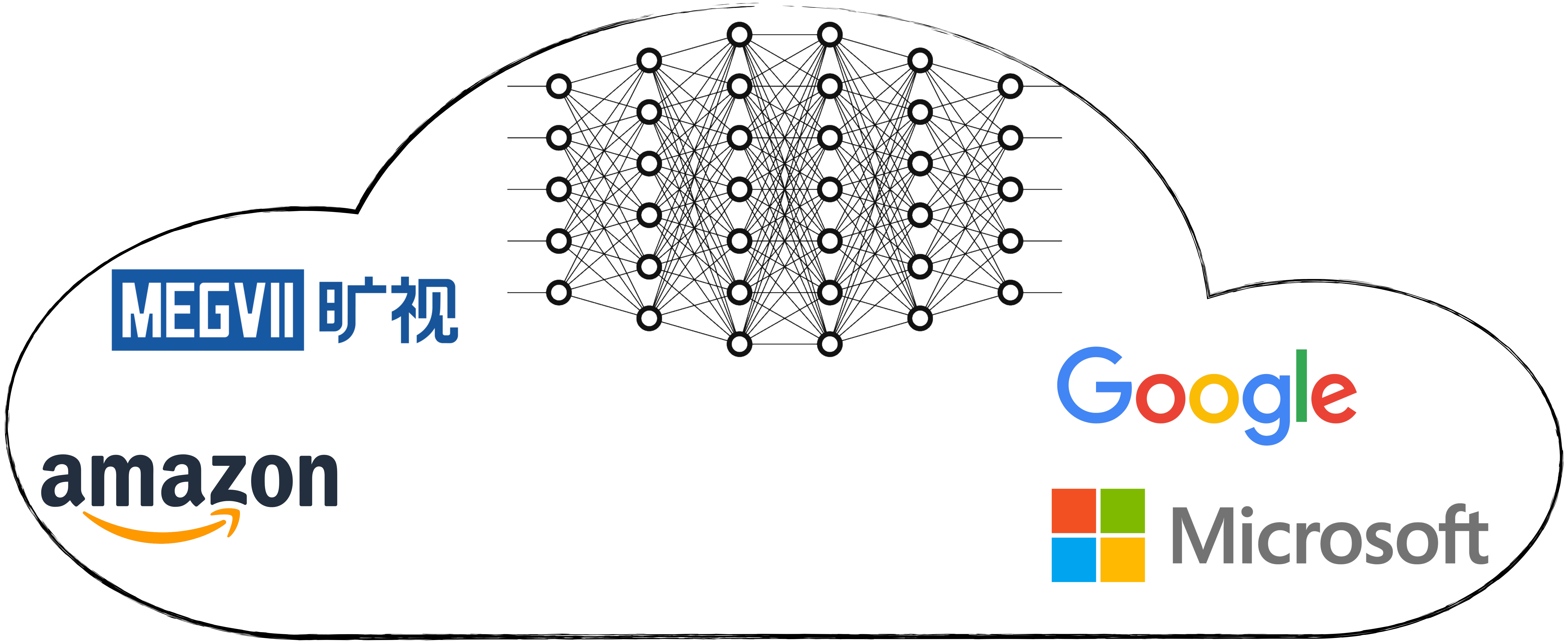
[1]Nanjing University [2]Cornell University

*Corresponding author: fengyuan.xu@nju.edu.cn

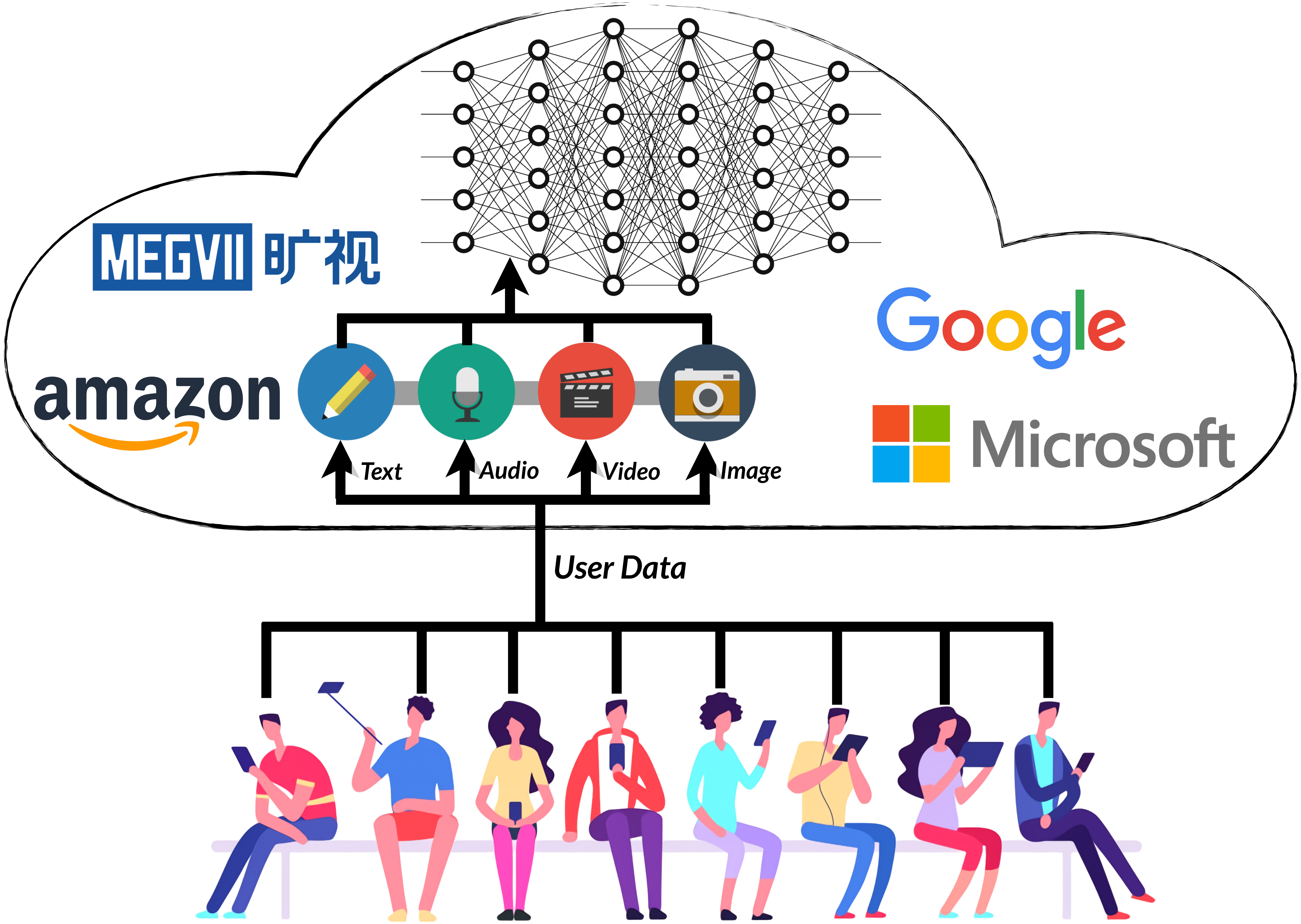

# Deep Learning Inference Service (*DLIS*) prospers.


# Deep Learning Inference Service (*DLIS*) prospers.

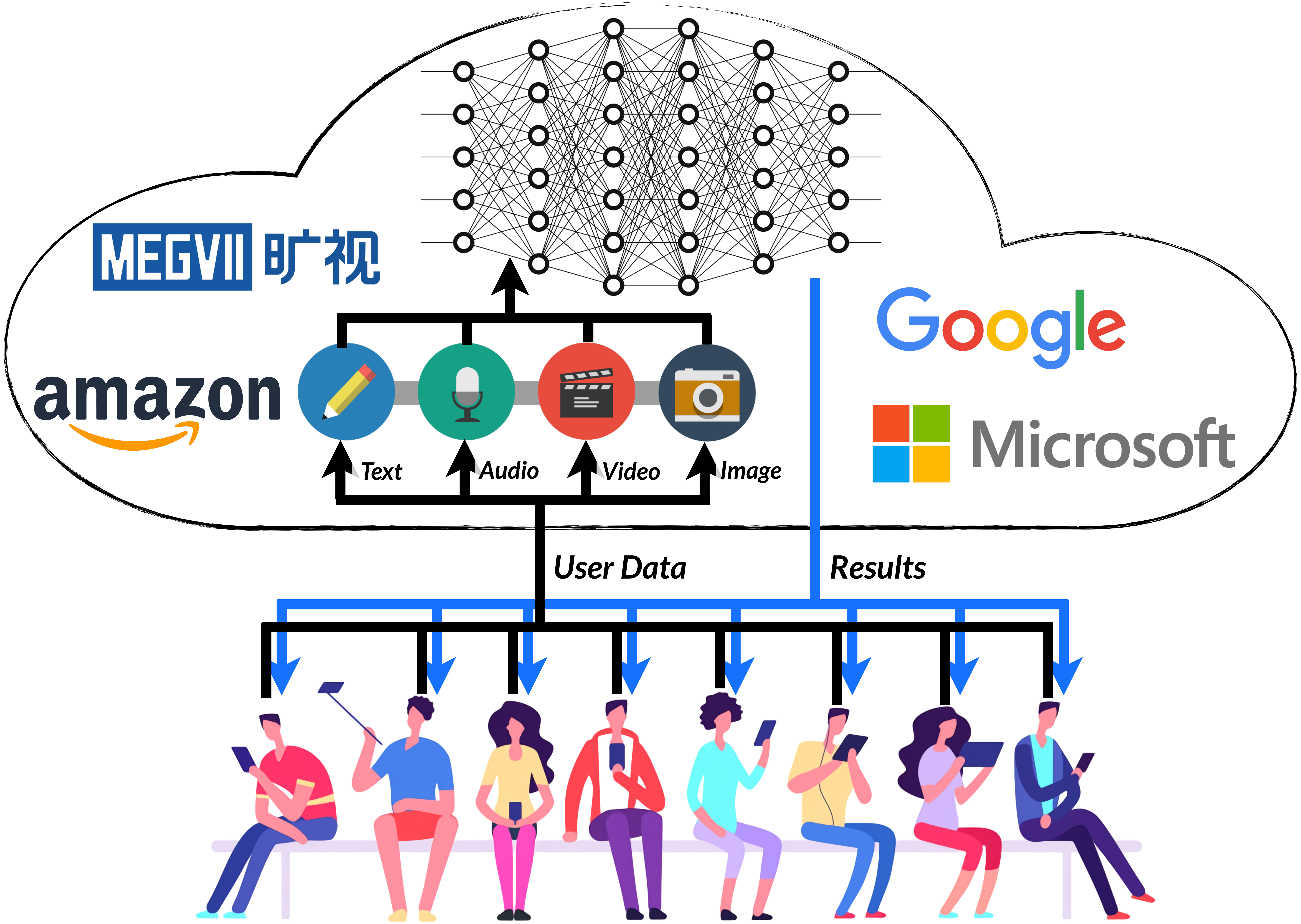
# DLIS Scenario




# DLIS Scenario

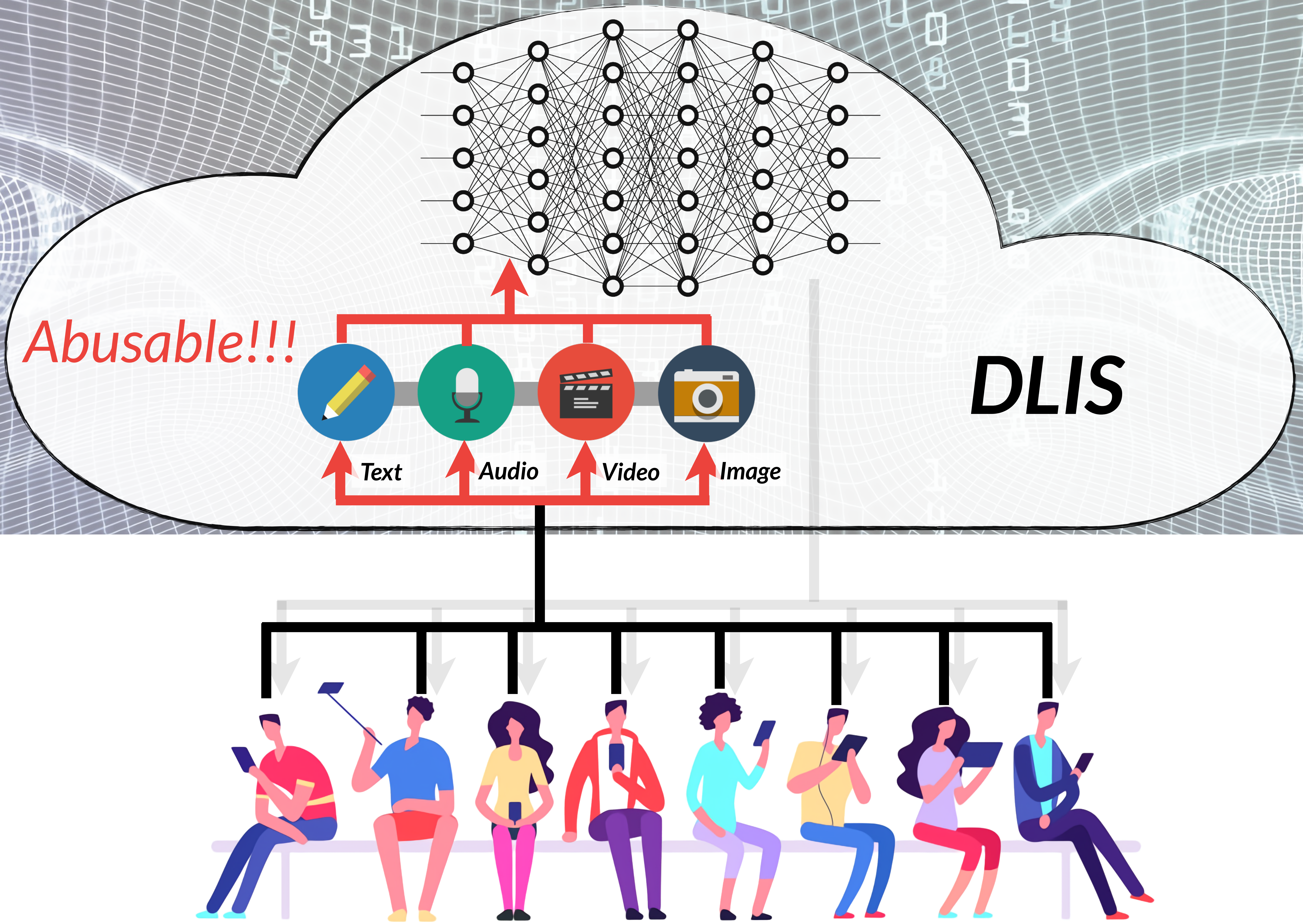


# DLIS Scenario

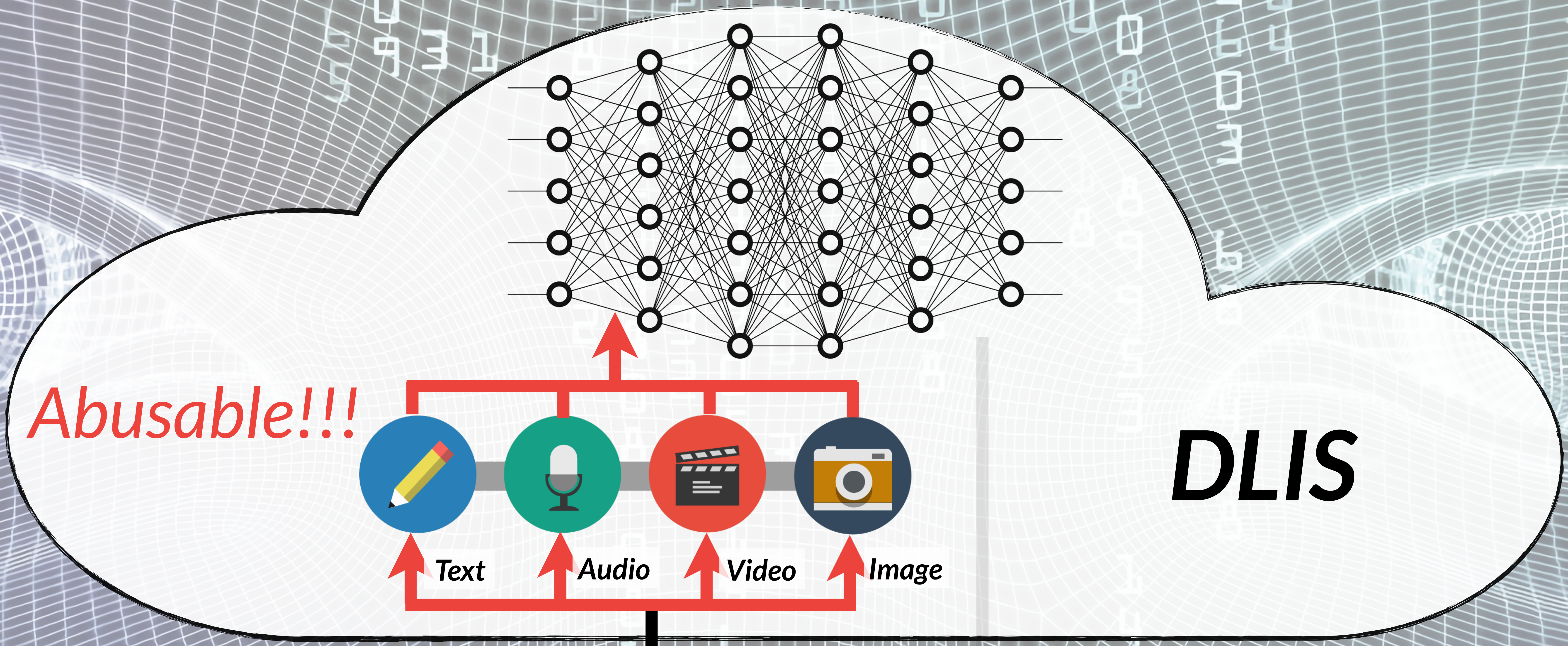


# Data abuse issue


# Data abuse issue



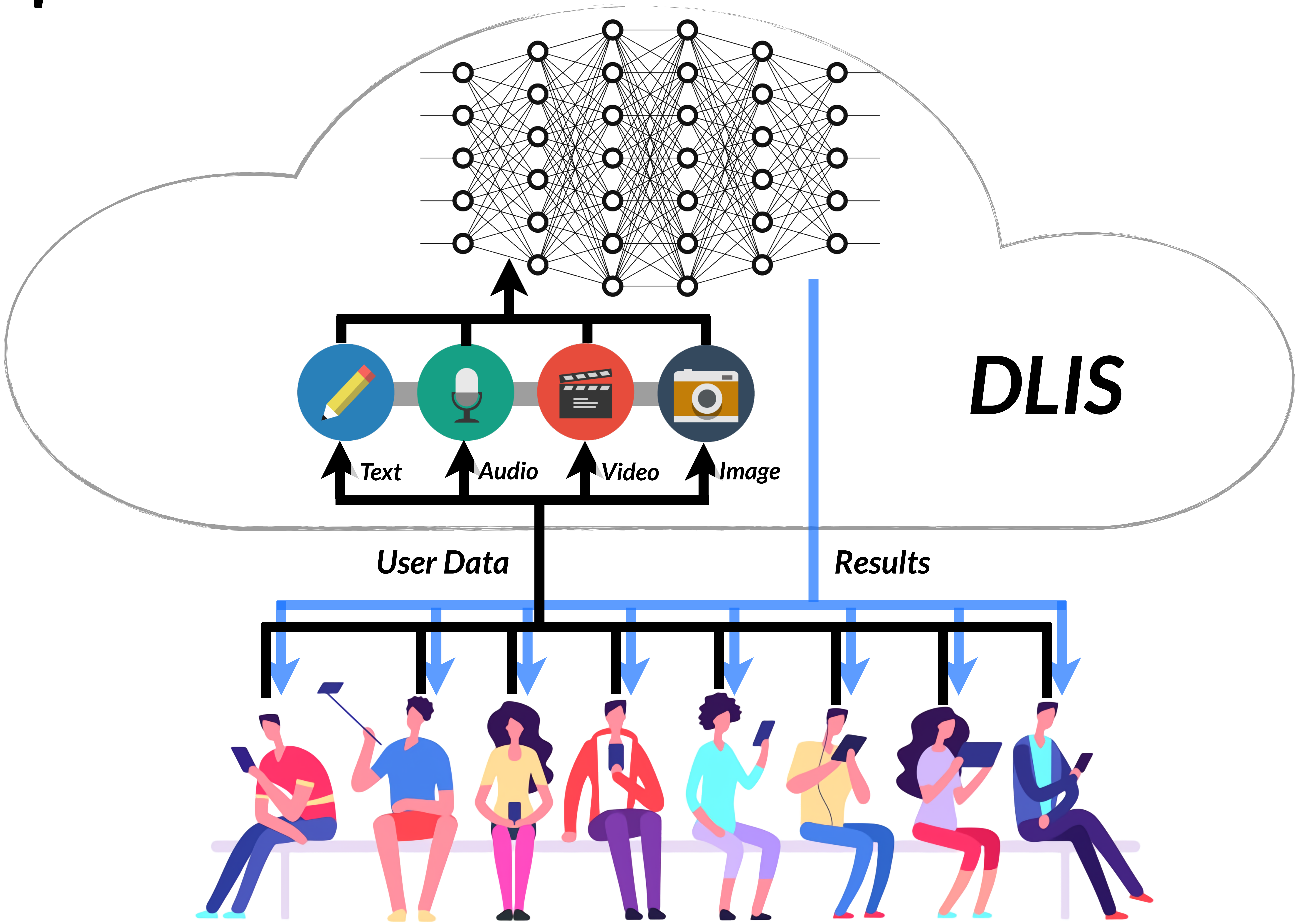
# Data abuse issue




**Data abuse is about the rights of data owners in the context of DLIS.**

1.Infer private info.
2.Train new models.

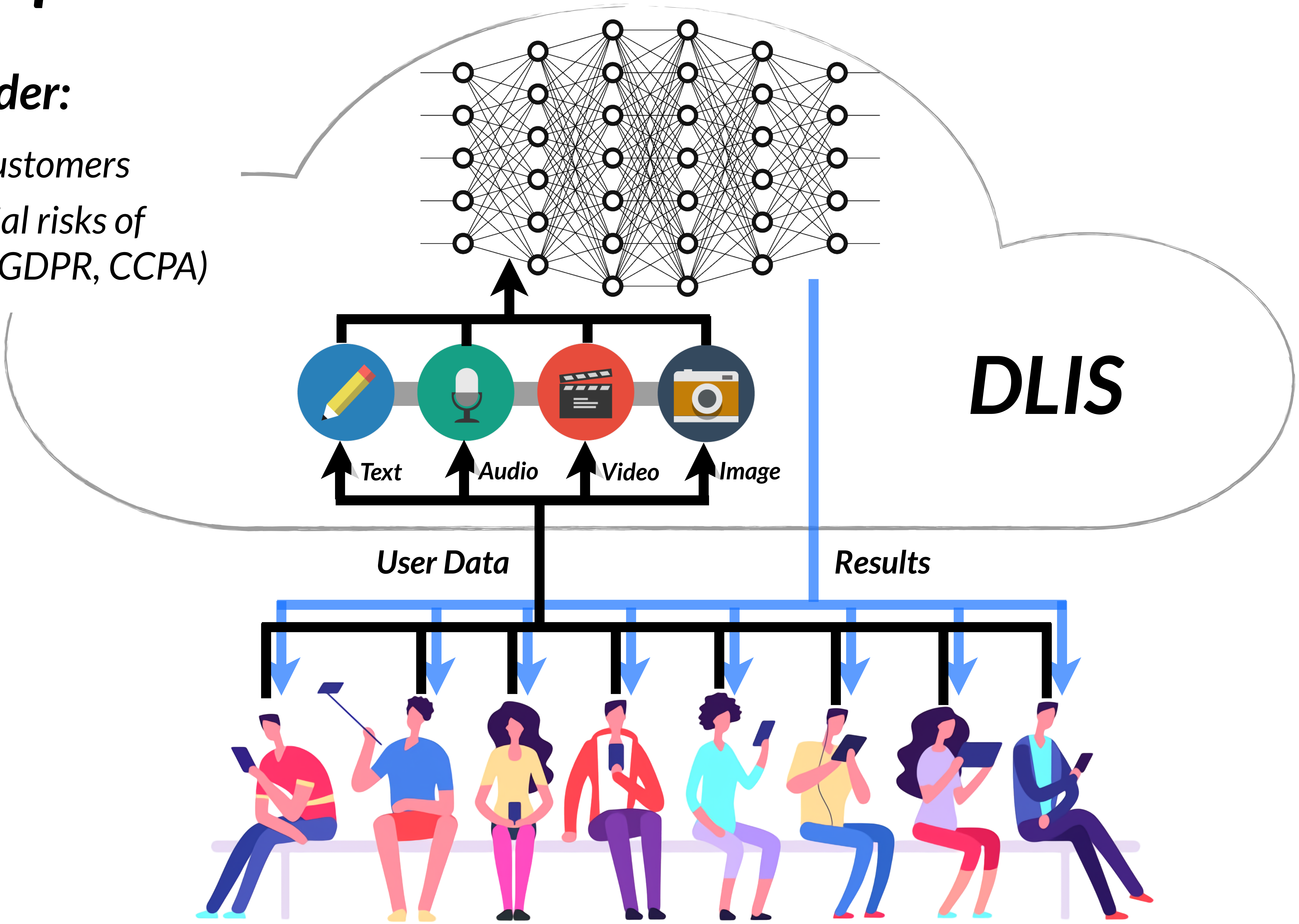
# Problem Requirements



# Problem Requirements

## Honest Provider:

- Attract more customers
- Reduce potential risks of violating laws (GDPR, CCPA)



# Problem Requirements

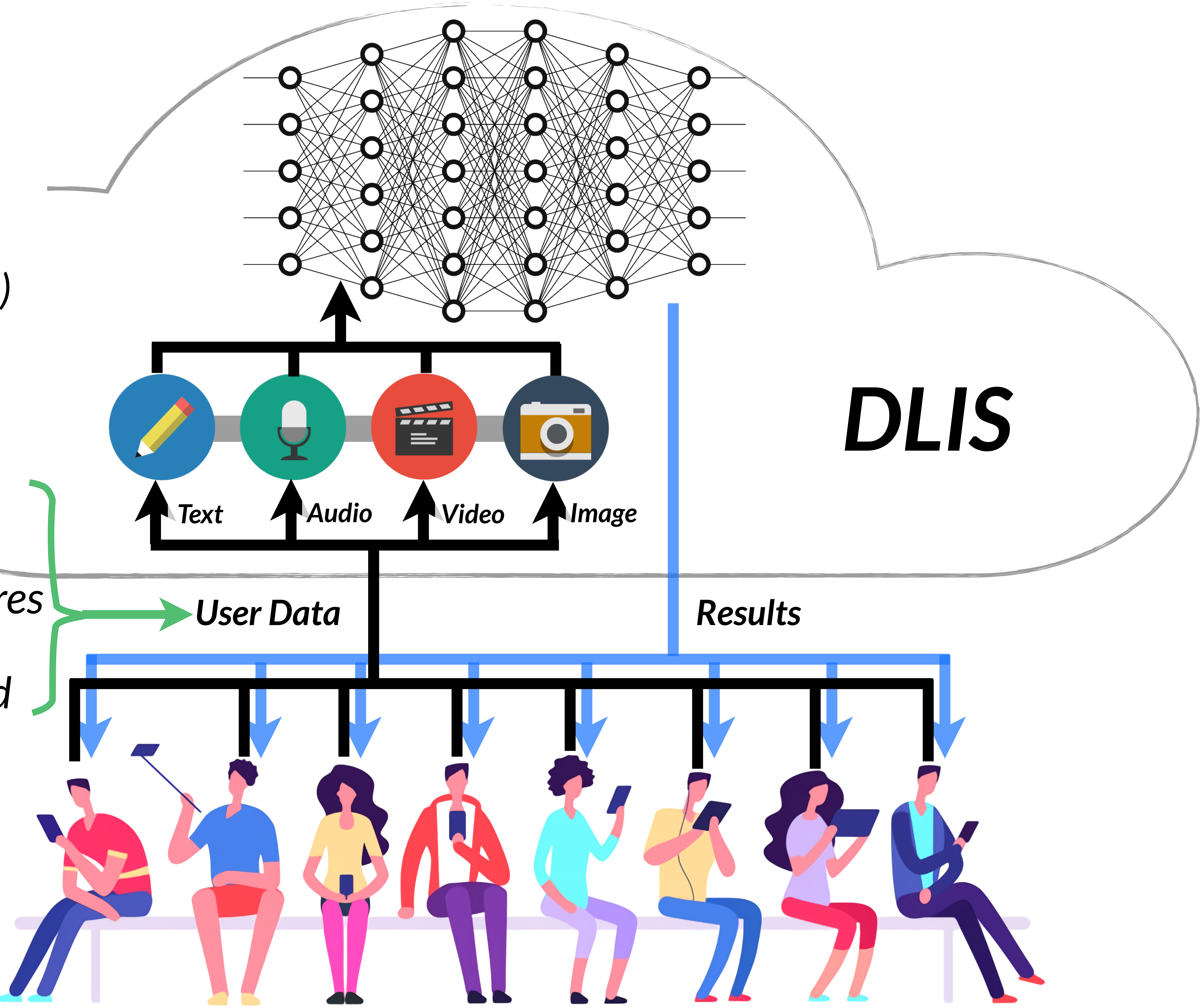
### Honest Provider:

- *Attract more customers*
- *Reduce potential risks of violating laws (GDPR, CCPA)*

*S1. Not visually recognizable*

*S2. Only retain necessary features*

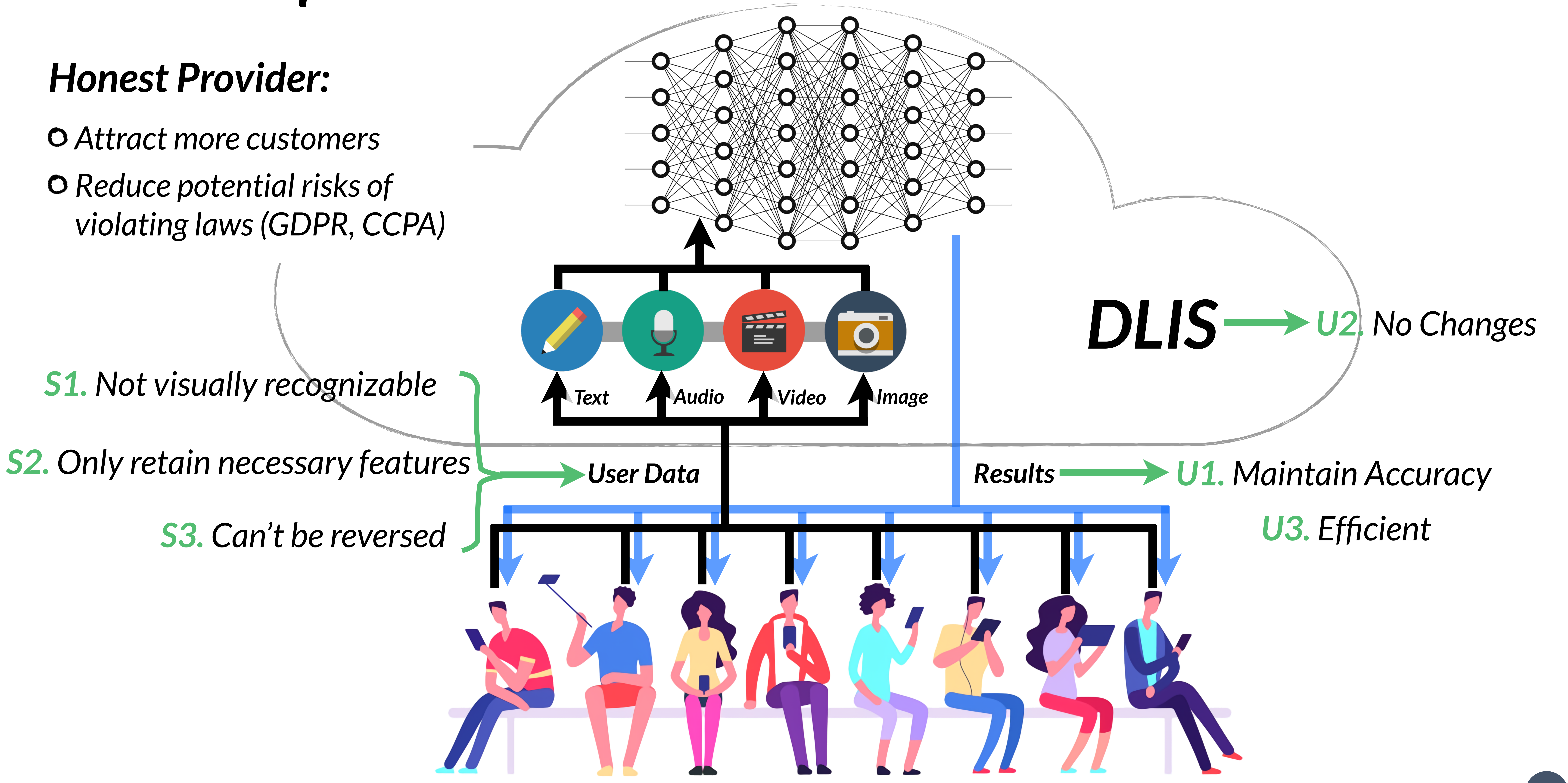
*S3. Can't be reversed*



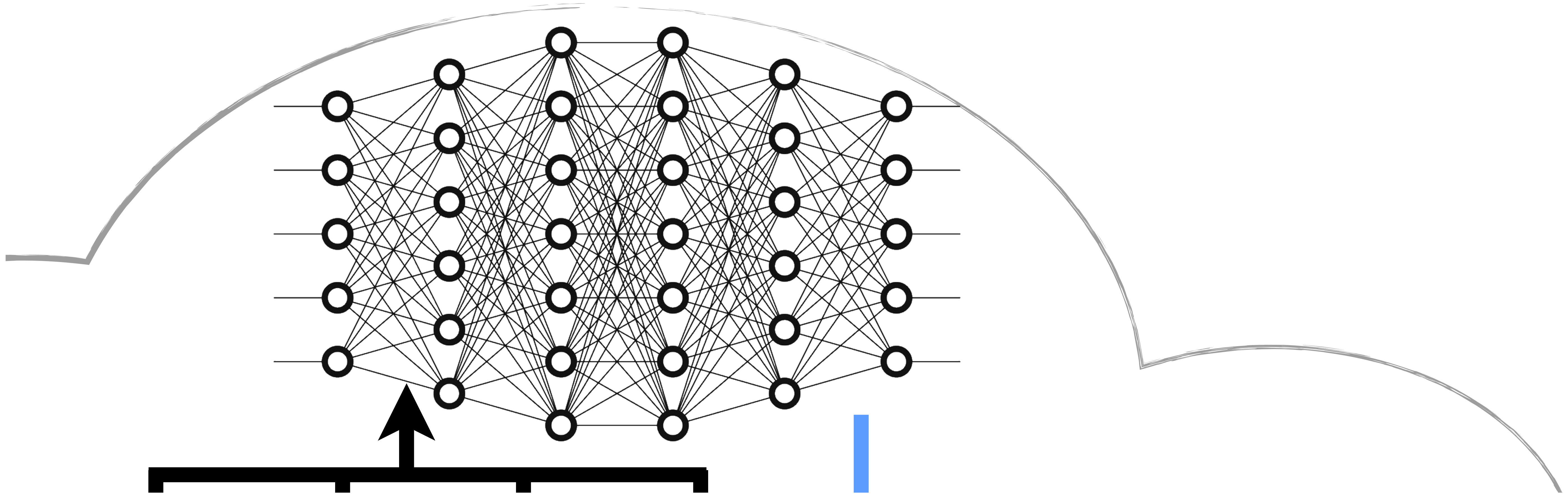
# Problem Requirements

### Honest Provider:

- Attract more customers
- Reduce potential risks of violating laws (GDPR, CCPA)



# Problem Requirements



**Honest Provider:**

- Attract more customers
- Reduce potential risks of violating laws (GDPR, CCPA)

**Security** ⟷ **Balance** ⟷ **Usability**

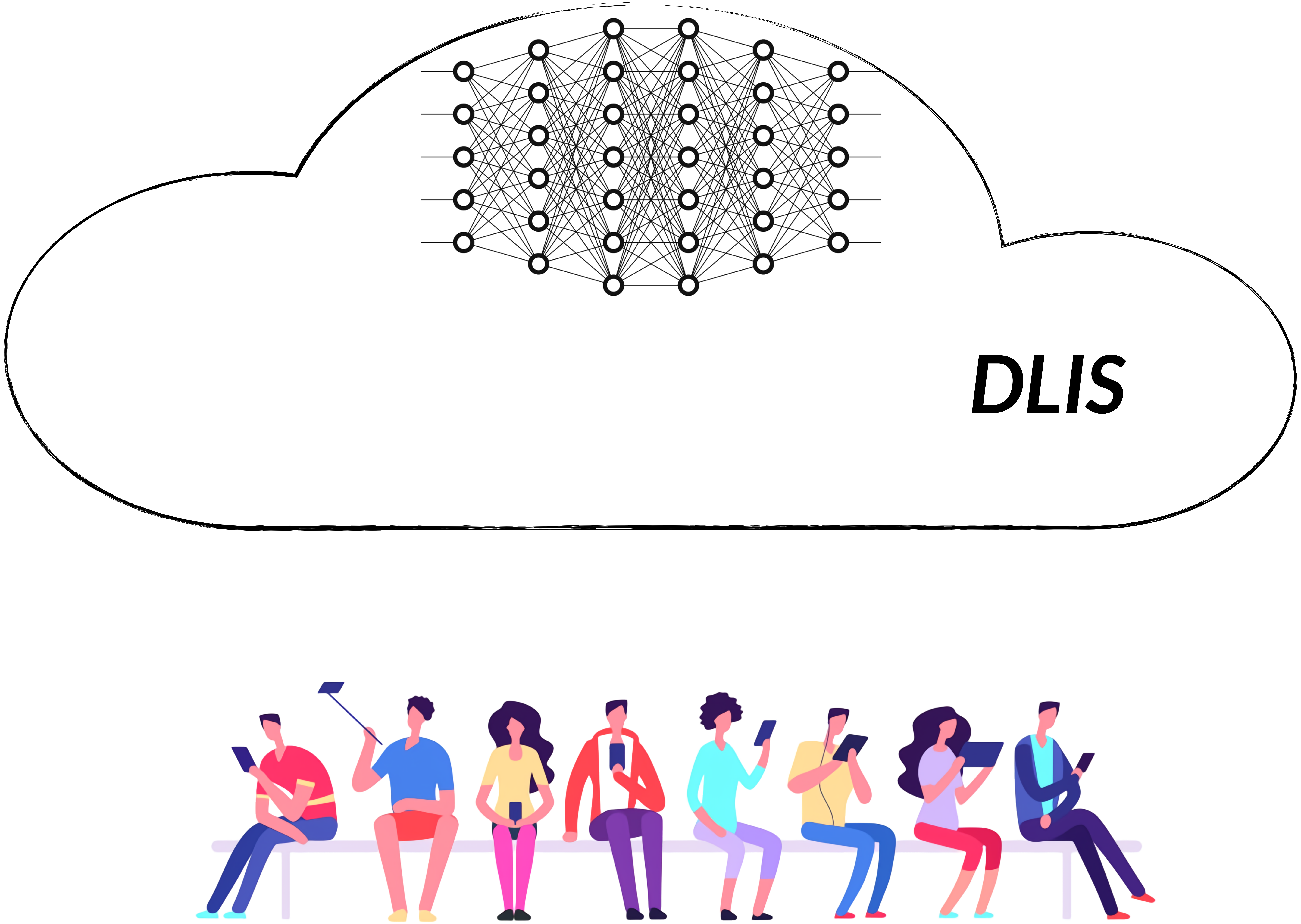
*Weak security solution*

- DP, MP, PAN

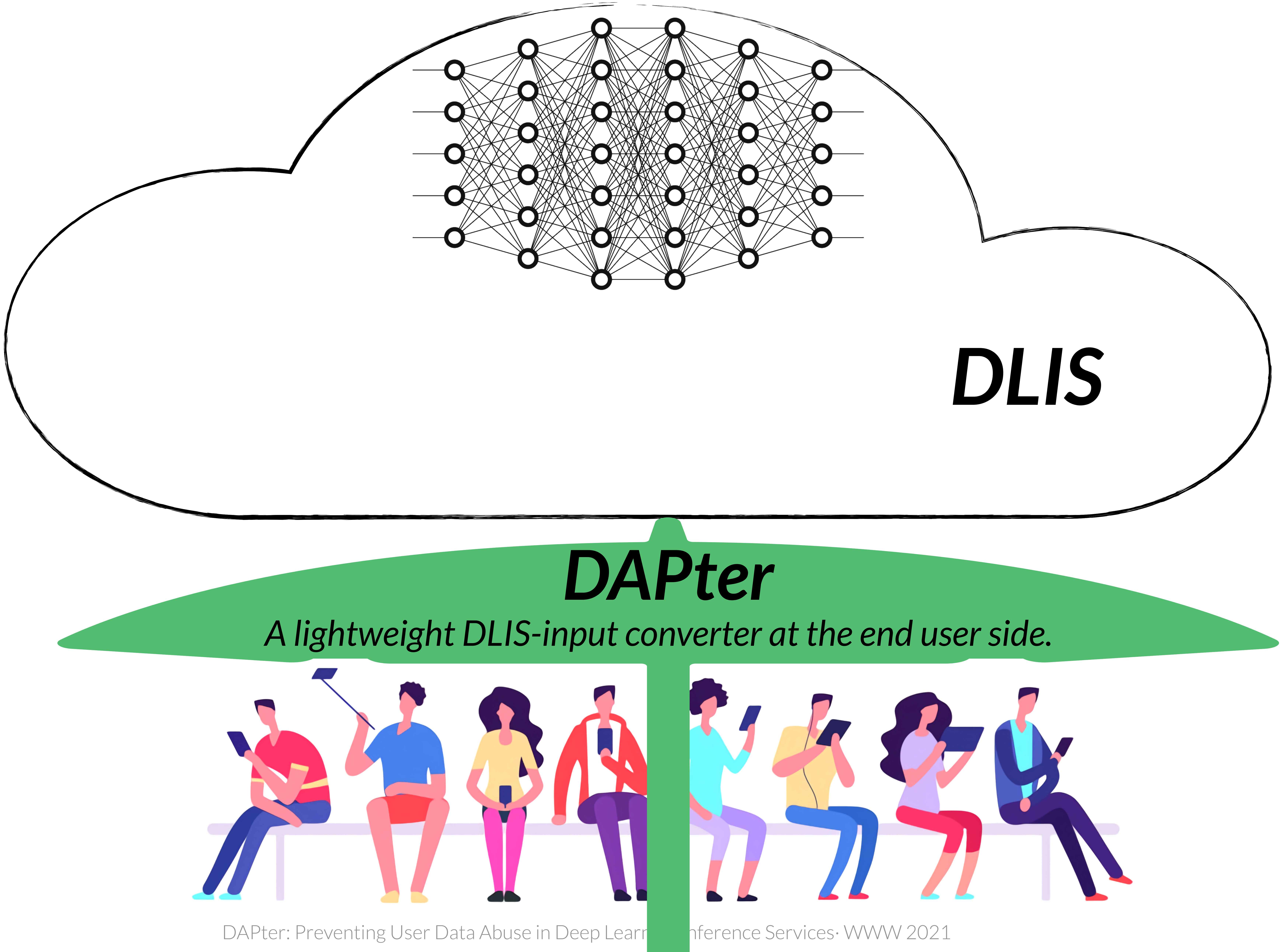
*Low usability solution*

- TEE, FHE

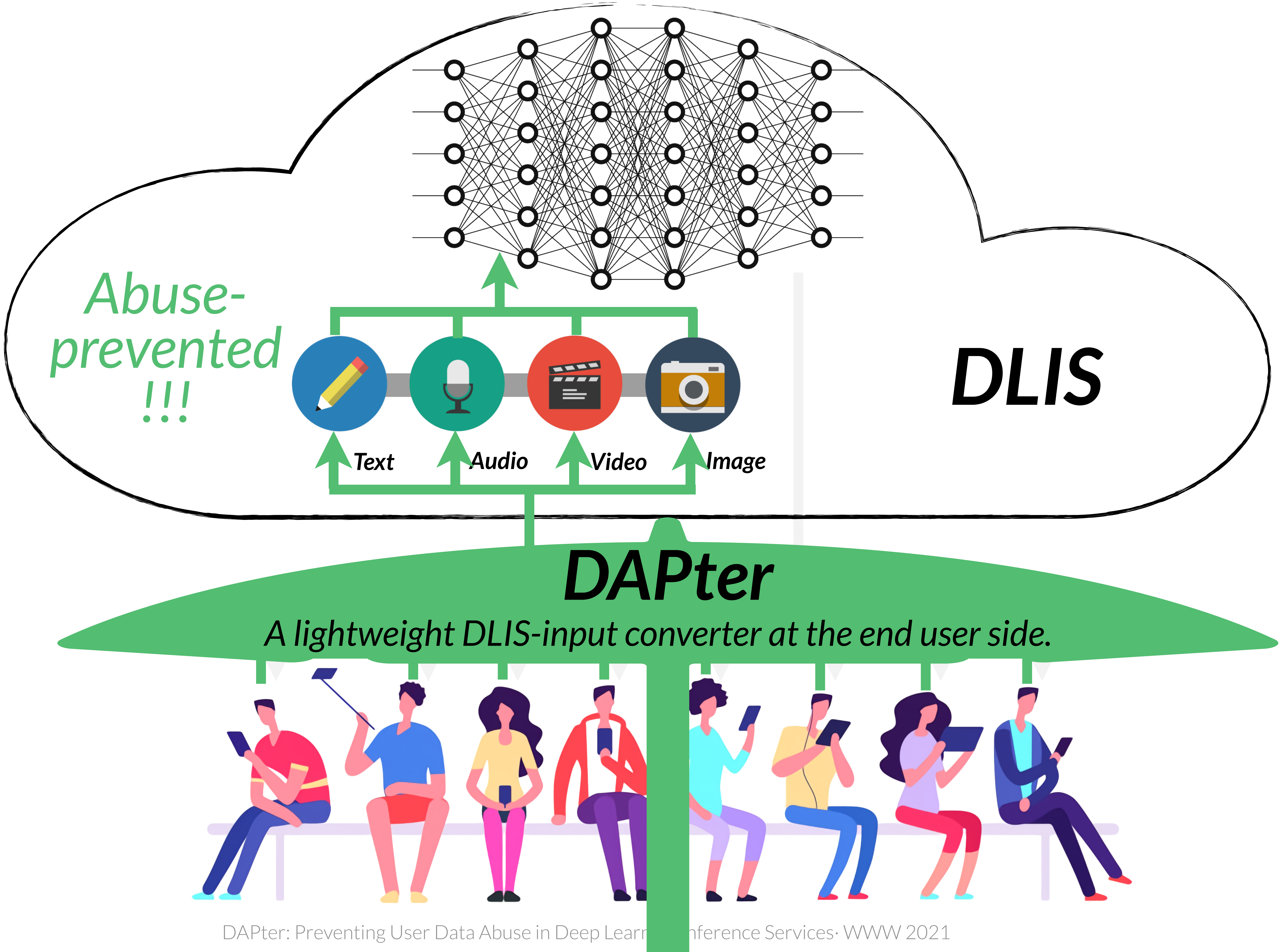
# Our solution DAPter



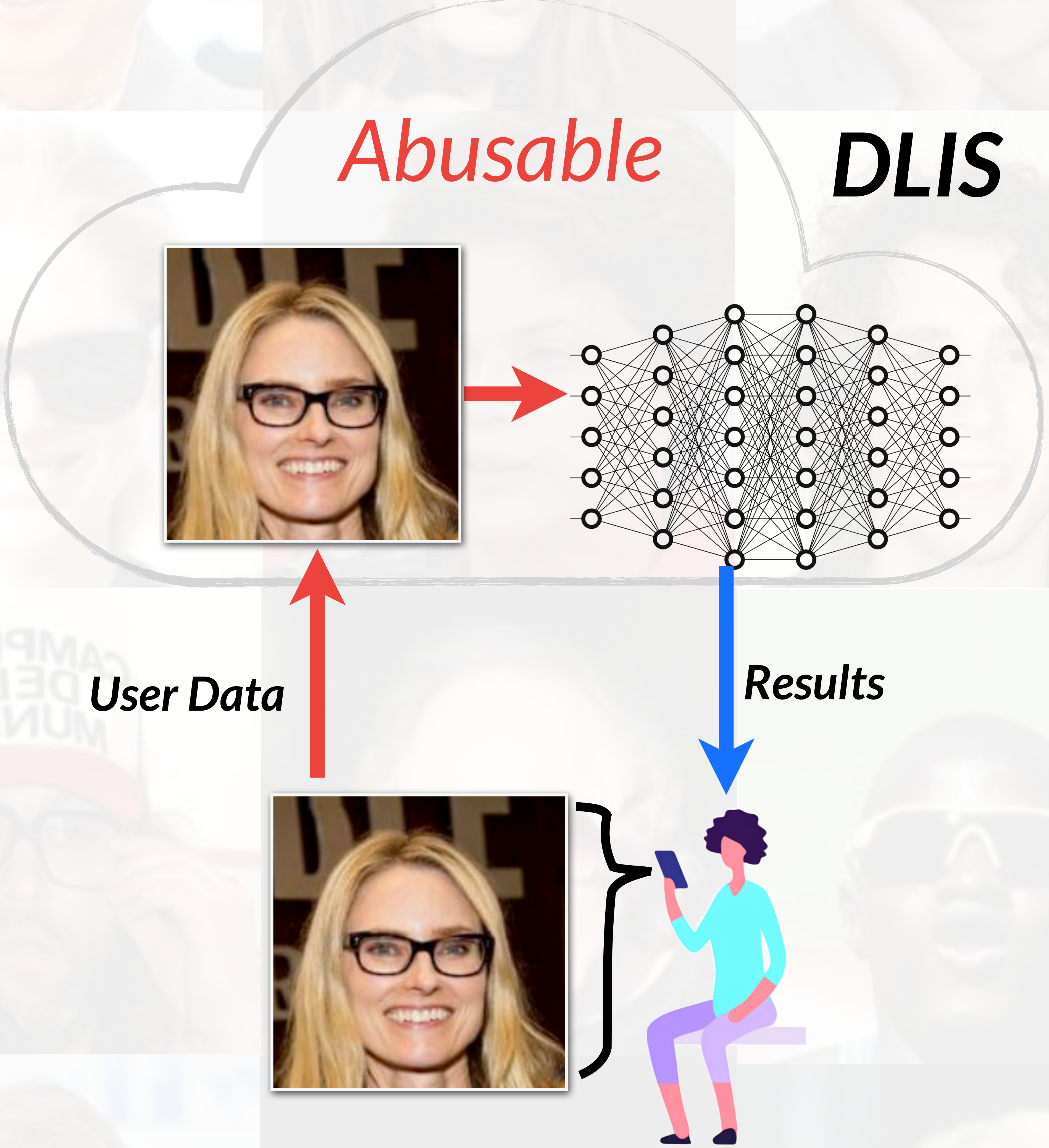
# *Our solution DAPter*



# Our solution DAPter

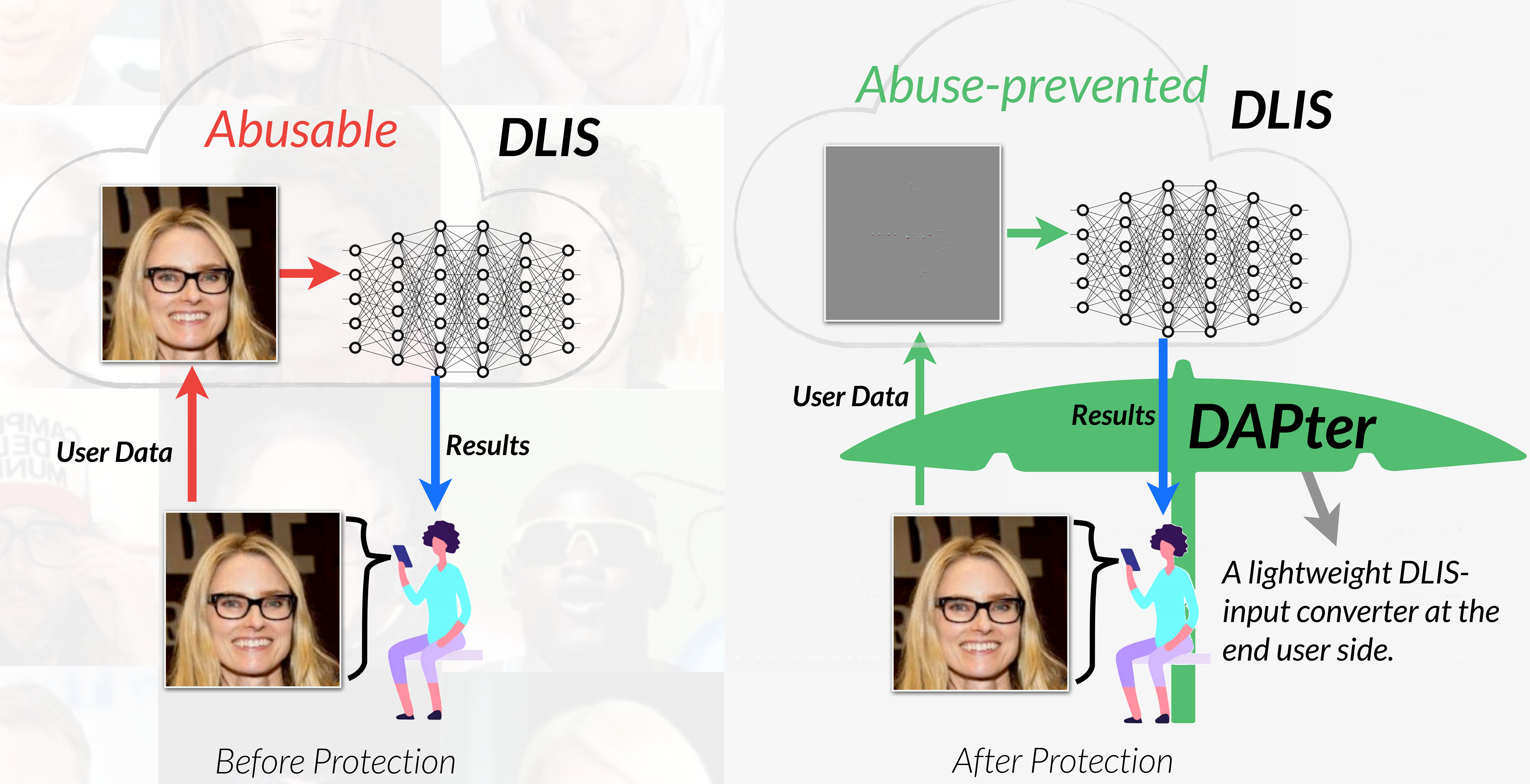


# DAPter Use Case



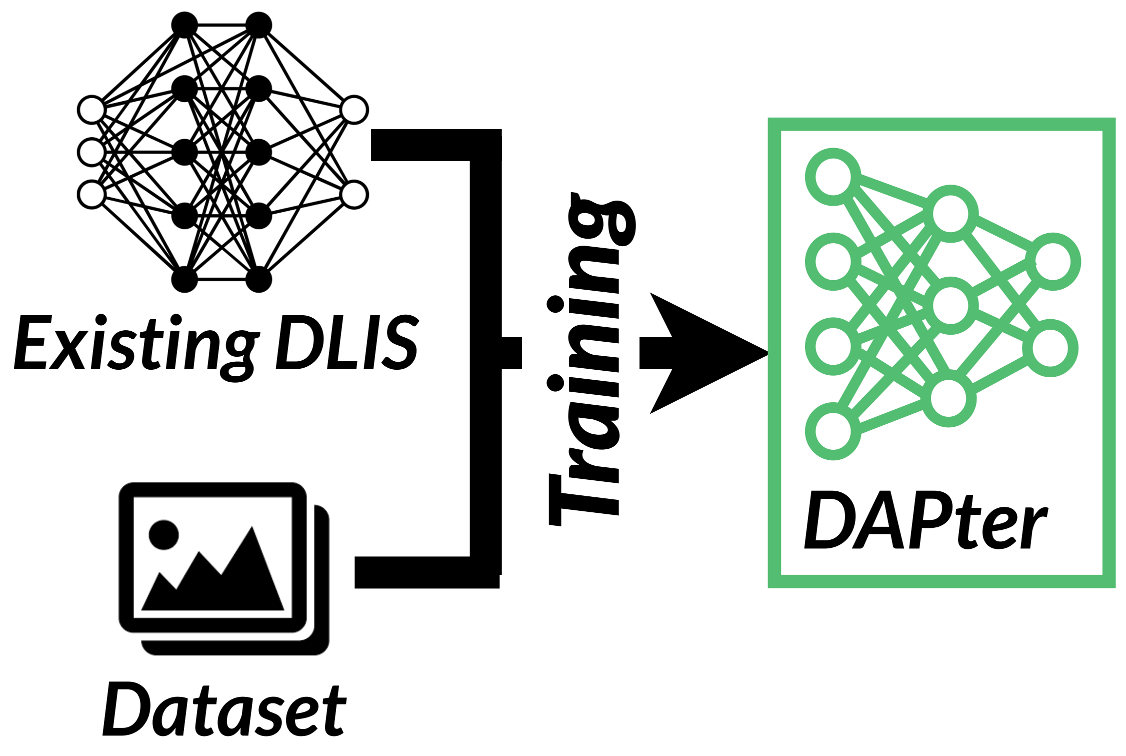

Before Protection

# DAPter Use Case



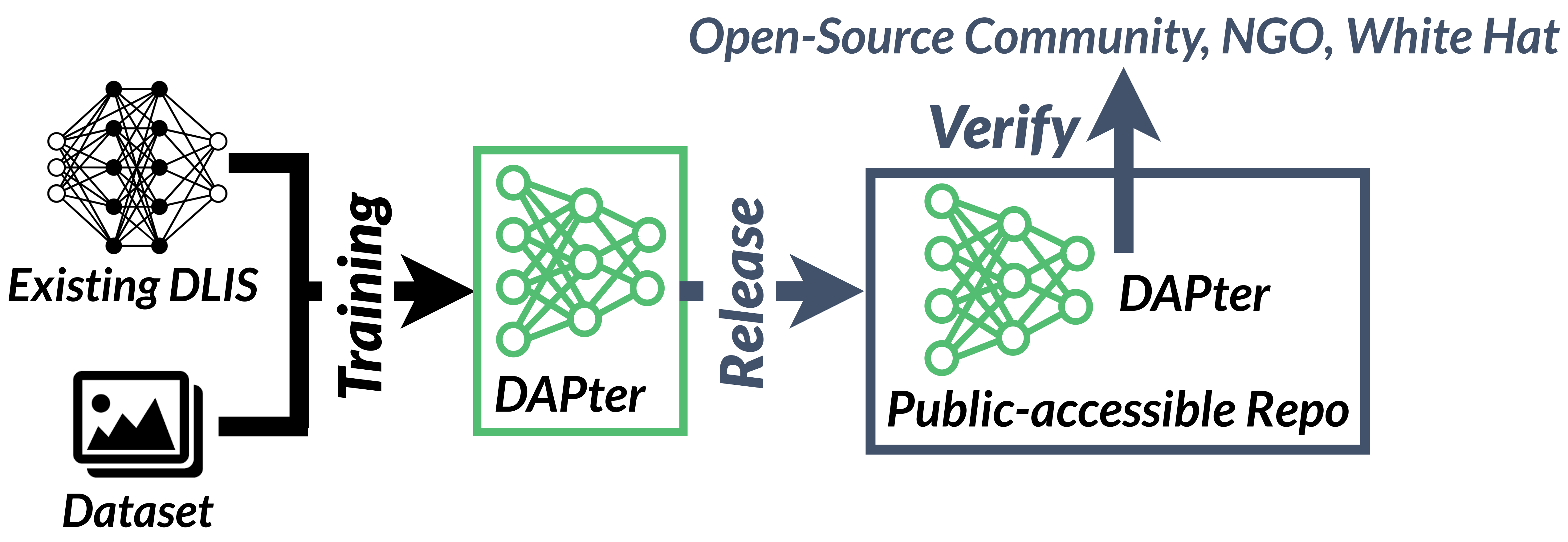
# Workflow

*A user-side **entropy reduction** approach to **prune information** not relevant to the target DLIS in user data.*



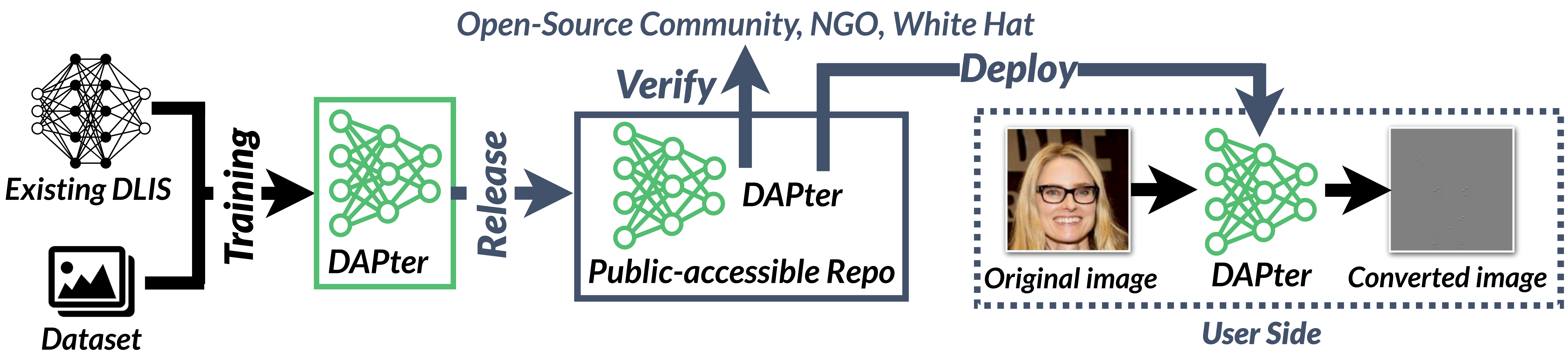
# Workflow

*A user-side* ***entropy reduction*** *approach to* ***prune information*** *not relevant to the target DLIS in user data.*



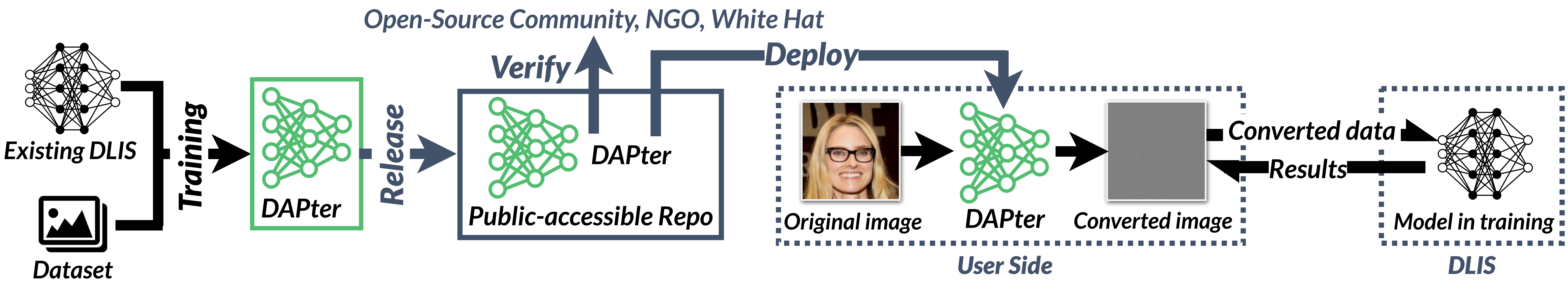
# Workflow

*A user-side **entropy reduction** approach to **prune information** not relevant to the target DLIS in user data.*



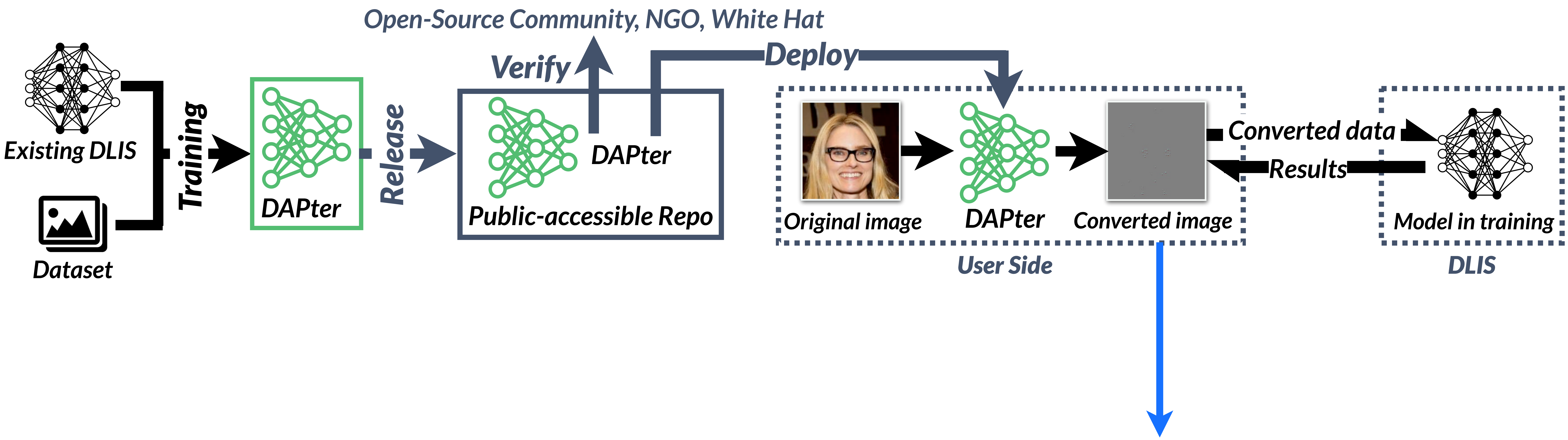
# Workflow

*A user-side **entropy reduction** approach to **prune information** not relevant to the target DLIS in user data.*



# Workflow

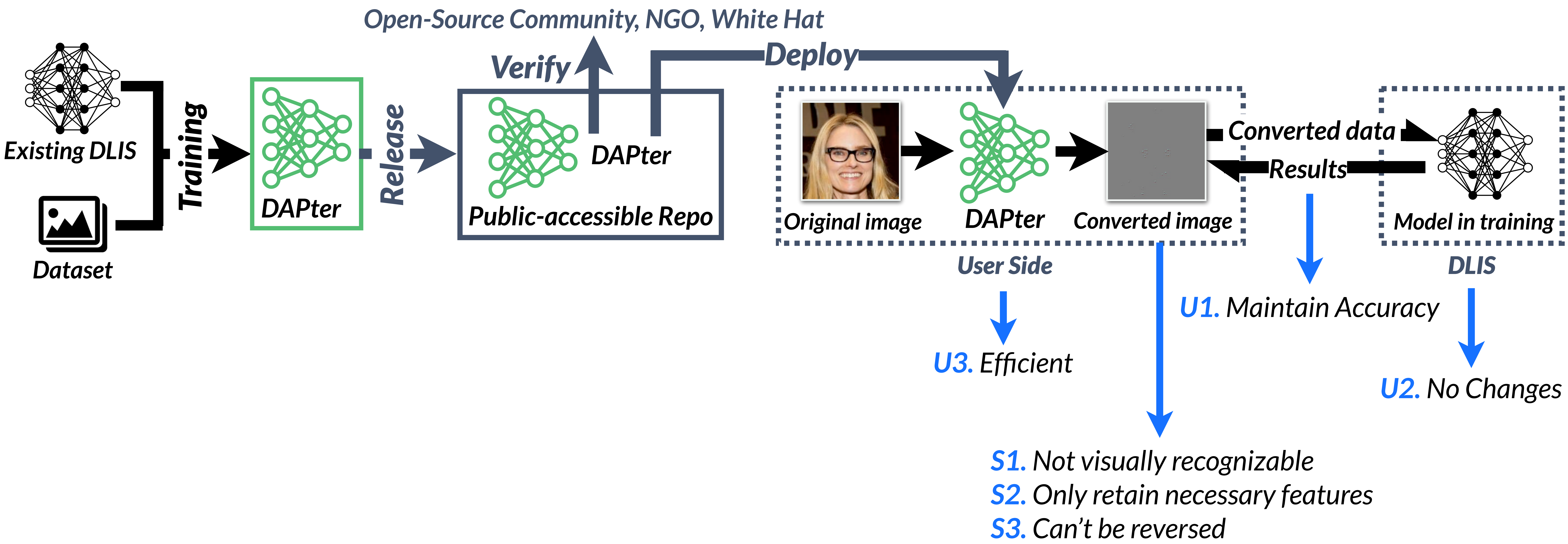
*A user-side **entropy reduction** approach to **prune information** not relevant to the target DLIS in user data.*




***S1.*** *Not visually recognizable*
***S2.*** *Only retain necessary features*
***S3.*** *Can't be reversed*

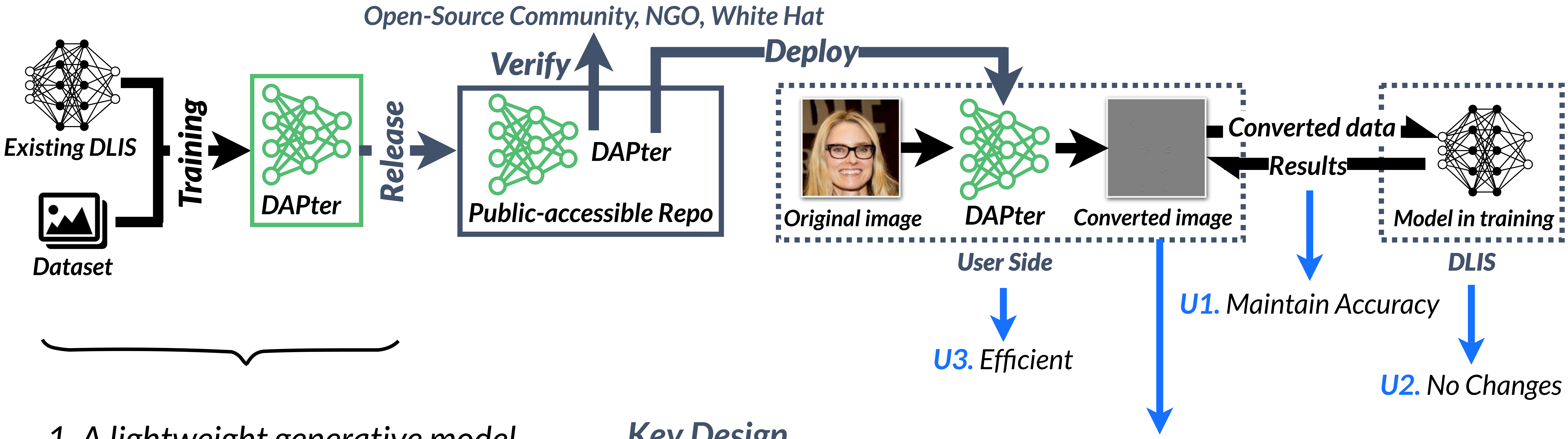
# Workflow

*A user-side **entropy reduction** approach to **prune information** not relevant to the target DLIS in user data.*



# Workflow

*A user-side **entropy reduction** approach to **prune information** not relevant to the target DLIS in user data.*

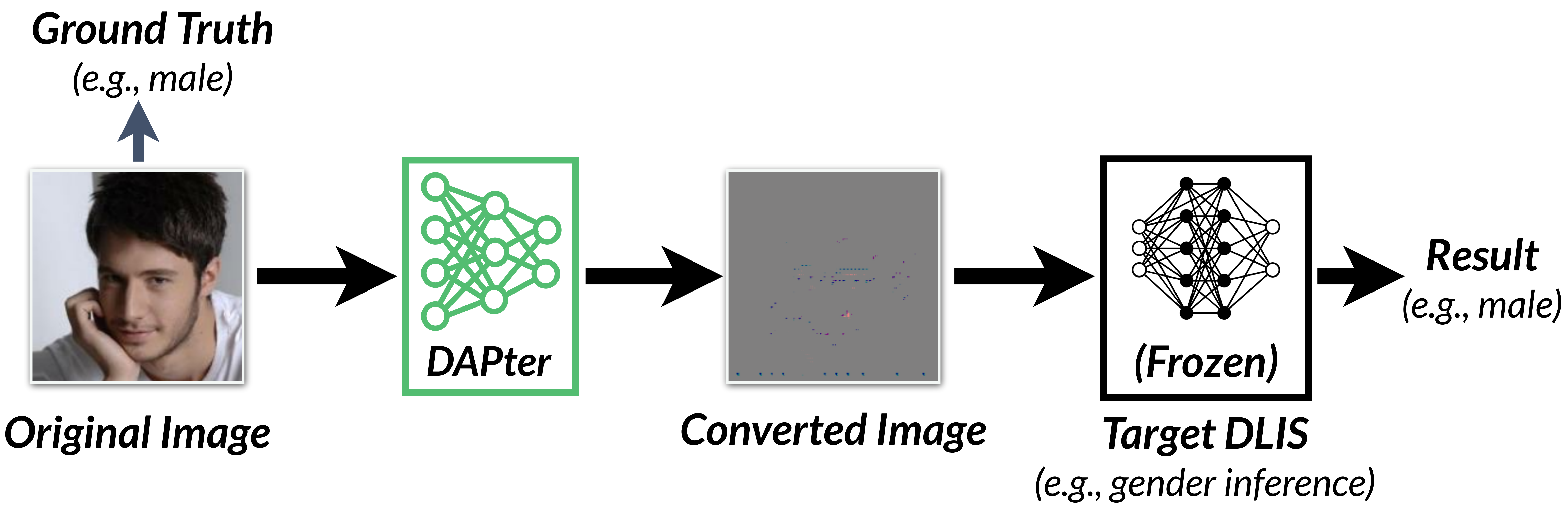



1. *A lightweight generative model*
2. *A data abuse prevention loss*

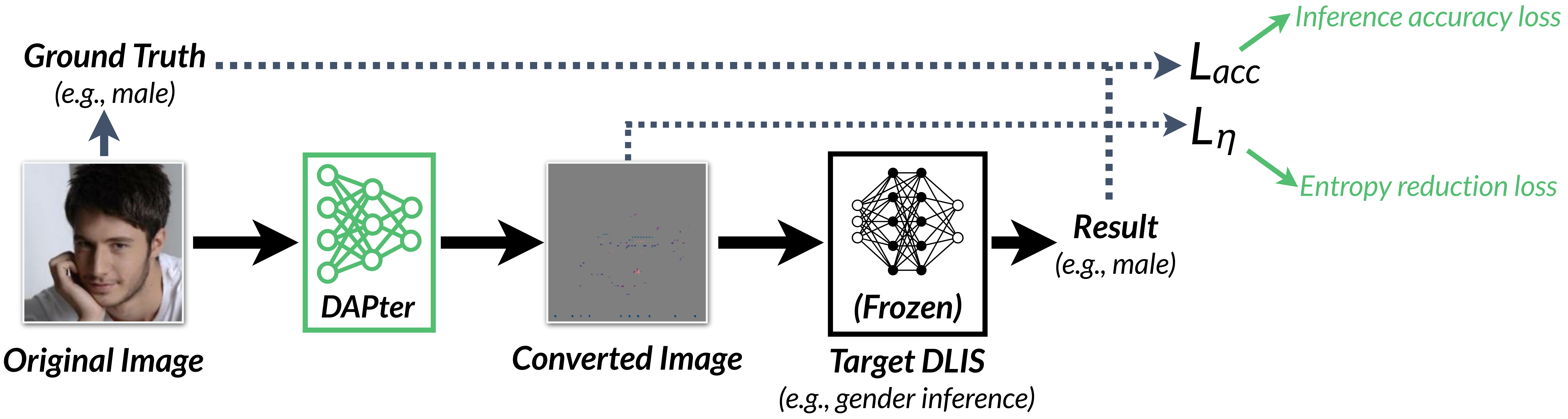
**Key Design**

*S1. Not visually recognizable*
*S2. Only retain necessary features*
*S3. Can't be reversed*

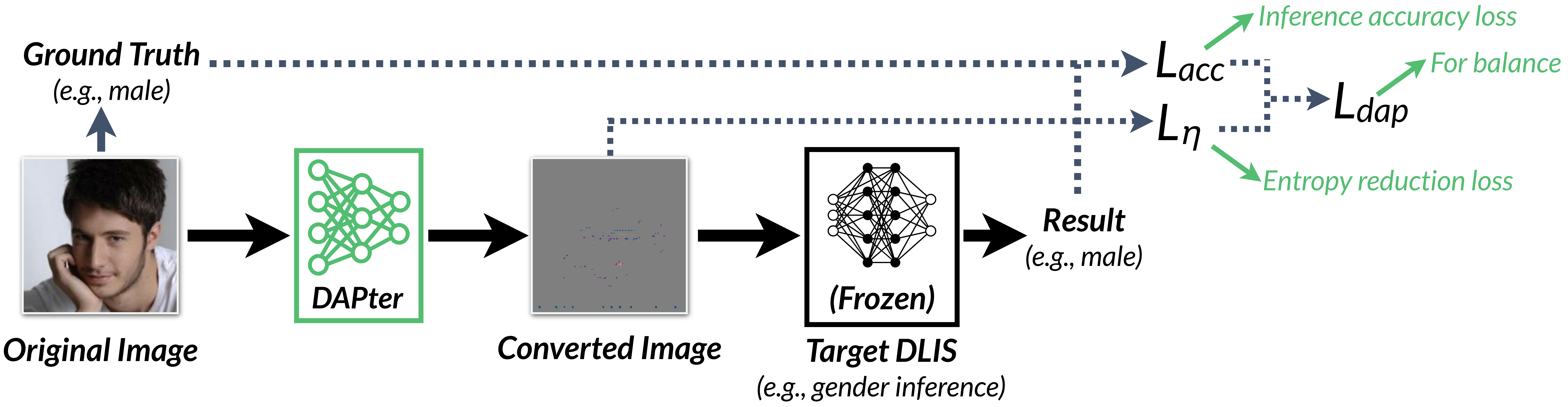
# Training Structure & Model Architecture



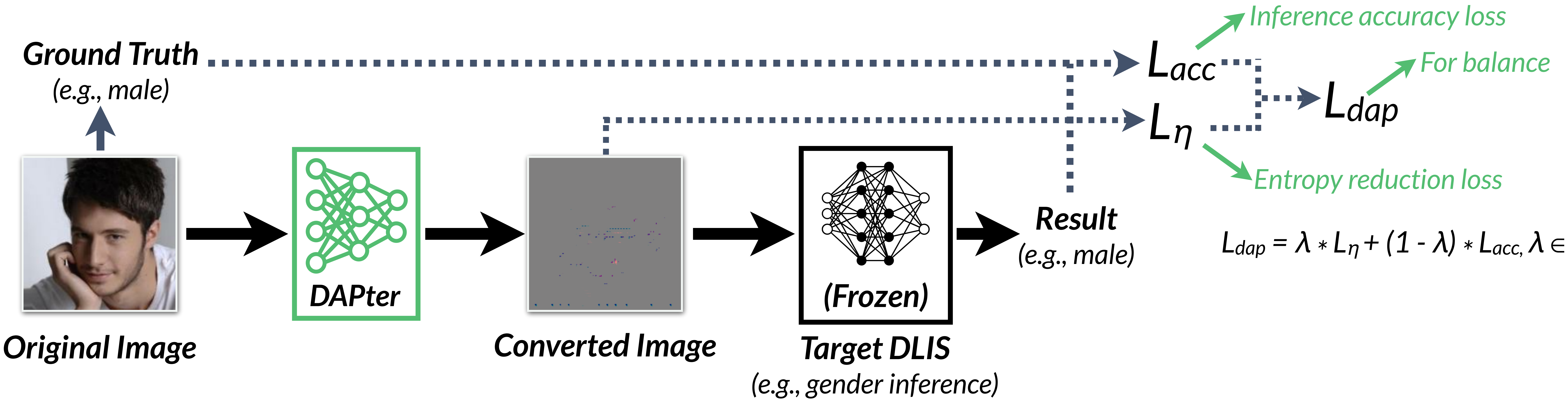
# Training Structure & Model Architecture



# Training Structure & Model Architecture

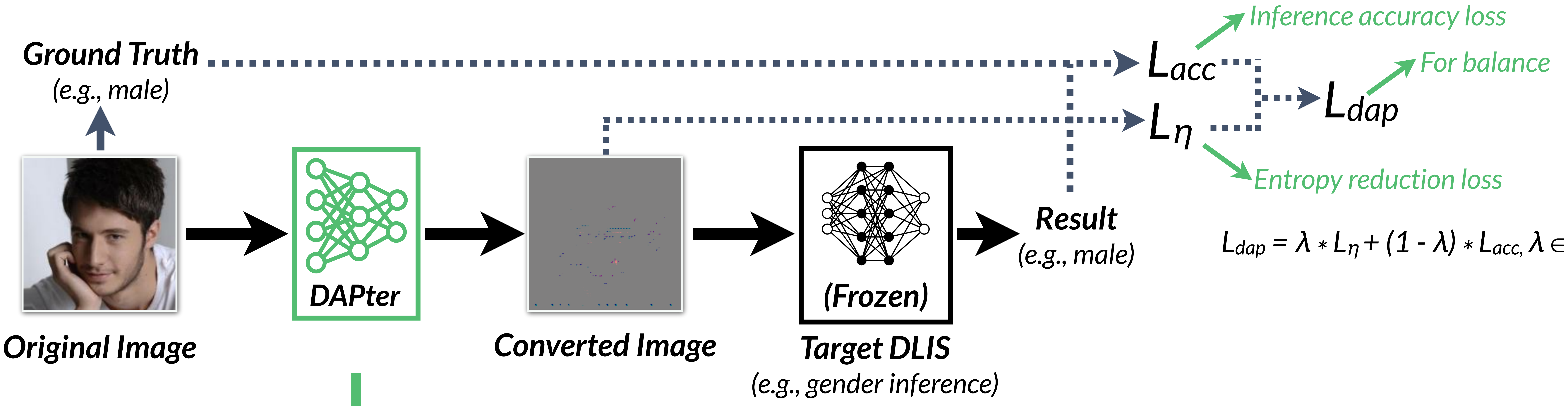


# Training Structure & Model Architecture

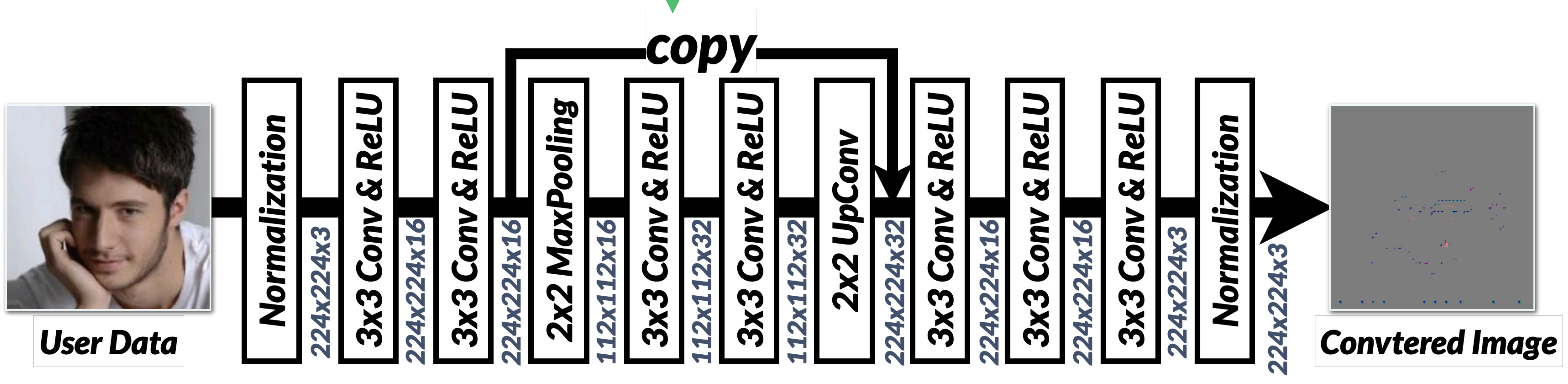



$L_{dap} = \lambda * L_{\eta} + (1 - \lambda) * L_{acc},\ \lambda \in (0, 1)$

# Training Structure & Model Architecture




$L_{dap} = \lambda * L_\eta + (1 - \lambda) * L_{acc}, \lambda \in (0, 1)$




Model Architecture

1. *A symmetrical U-Net like architecture.*
2. *Input and output are of the same size.*
3. *"Copy" connection captures the high-level semantic info and low-level spatial info.*

# *Data Abuse Prevention Loss*

*Minimize the piece of pixel-wise entroy that contributes little to the high-level features.*

$$L_{dap} = \lambda * L_{\eta} + (1 - \lambda) * L_{acc}, \lambda \in (0, 1)$$

# Data Abuse Prevention Loss

*Minimize the piece of pixel-wise entroy that contributes little to the high-level features.*

$$L_{dap} = \lambda * L_{\eta} + (1 - \lambda) * L_{acc}, \lambda \in (0, 1)$$

$L_{acc}$ *measures the inference accuracy of the target DLIS.*

$L_{\eta}$ *measures the pixel-wise entropy* $(H_I = -\sum_i p_i \log p_i)$ *in input data.* $p_i$ *is the occurrence possiblility of i.*

# Data Abuse Prevention Loss

*Minimize the piece of pixel-wise entroy that contributes little to the high-level features.*

$$L_{dap} = \lambda * L_{\eta} + (1 - \lambda) * L_{acc},\ \lambda \in (0, 1)$$

$L_{acc}$ *measures the inference accuracy of the target DLIS.*

$L_{\eta}$ *measures the pixel-wise entropy* $(H_I = -\sum_i p_i \log p_i)$ *in input data.* $p_i$ *is the occurrence possiblility of i.*

***Support statement*.*** *By enlarging the occurrence possibility of a specific pixel value, the upper bound of an image's pixel-wise entropy can be reduced.*

**Proof can be found in our paper*

# Data Abuse Prevention Loss

*Minimize the piece of pixel-wise entroy that contributes little to the high-level features.*

$$L_{dap} = \lambda * L_{\eta} + (1 - \lambda) * L_{acc},\ \lambda \in (0, 1)$$

$L_{acc}$ *measures the inference accuracy of the target DLIS.*

$L_{\eta}$ *measures the pixel-wise entropy* $(H_I = -\sum_i p_i \log p_i)$ *in input data.* $p_i$ *is the occurrence possiblility of i.*

***Support statement*.*** *By enlarging the occurrence possibility of a specific pixel value, the upper bound of an image's pixel-wise entropy can be reduced.*

$$L_{\eta} = \sum_I \eta(I, I_{ref})$$

***η*** *is L1 norm;* ***I*** *is the converted image;* $\mathbf{I_{ref}}$ *is the reference image with each pixel equaling to (R128, G128, B128).*

**Proof can be found in our paper*

# Heyperparameter λ Exploration

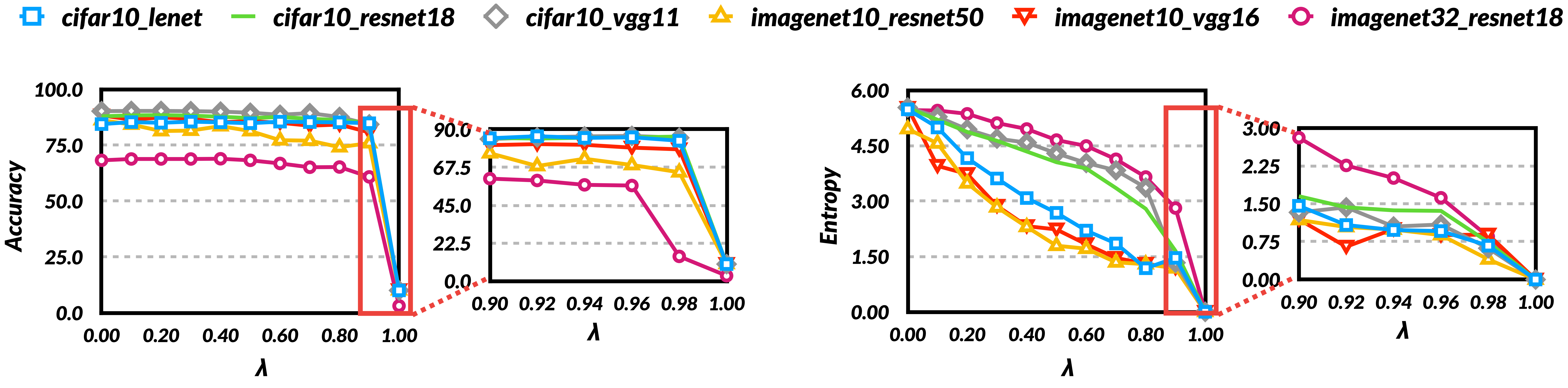
$$L_{dap} = \lambda * L_{\eta} + (1 - \lambda) * L_{acc},\ \lambda \in (0, 1)$$

*A larger λ lets DAPter remove more entropy but leads to a low DLIS accuracy.*

# Heyperparameter λ Exploration

$$L_{dap} = \lambda * L_{\eta} + (1 - \lambda) * L_{acc},\ \lambda \in (0, 1)$$

*A larger λ lets DAPter remove more entropy but leads to a low DLIS accuracy.*



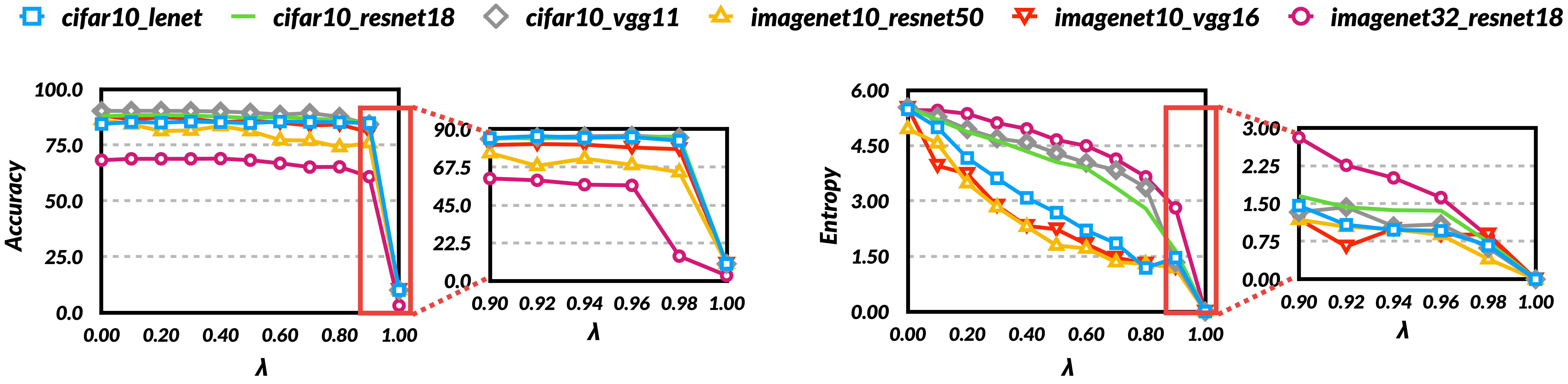

**(a) Relationship between λ and Accuracy**

**(b) Relationship between λ and Entropy**

# *Heyperparameter λ Exploration*

$$L_{dap} = \lambda * L_{\eta} + (1 - \lambda) * L_{acc}, \lambda \in (0, 1)$$

*A larger λ lets DAPter remove more entropy but leads to a low DLIS accuracy.*




***(a) Relationship between λ and Accuracy***

***(b) Relationship between λ and Entropy***

*λ = 0.9 is a sweet point to balance security and usability.*

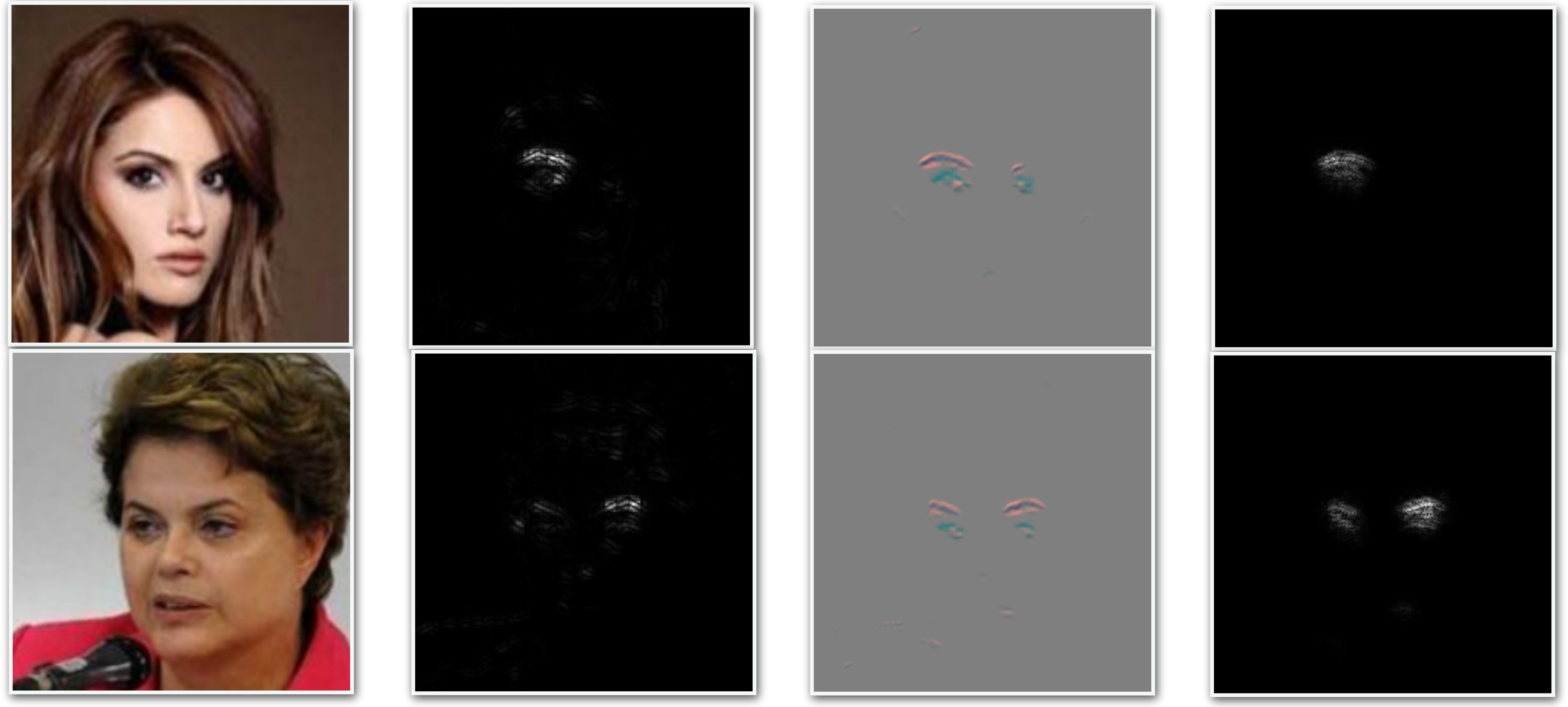
# Conversion Quality

*To show that DAPter can remove the unnecessary features and retain the useful features, we generate saliency map (SM) to measure which part of the input supports the DLIS through Grad-CAM.*

# Conversion Quality

*To show that DAPter can remove the unnecessary features and retain the useful features, we generate saliency map (SM) to measure which part of the input supports the DLIS through Grad-CAM.*

*Results are visualize below. From left to right is original image, sm of DLIS, protected image, sm of DAPter-enabled DLIS.*



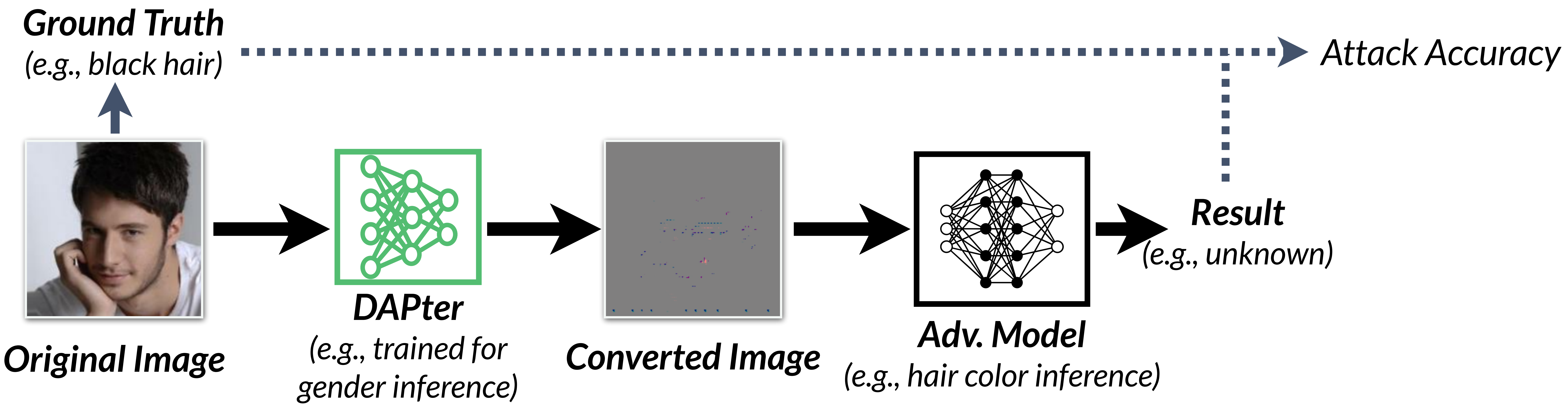
*(a) Arched Eyebrow Inference*

*(b) Wearing Glasses Inference*

*(c) Gender Inference*

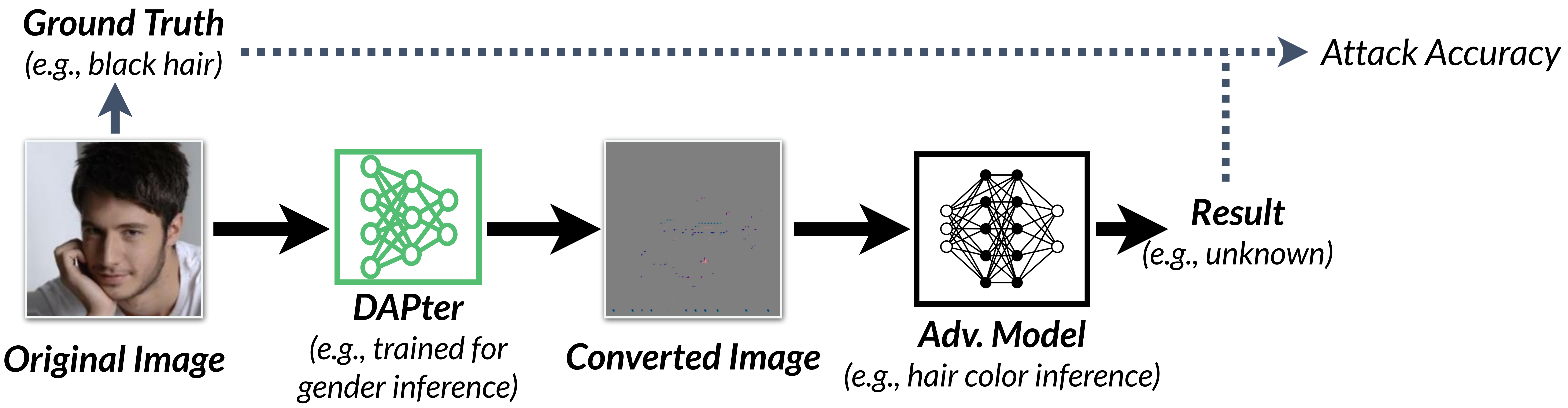
# *Security - Auto Recognition Attack*

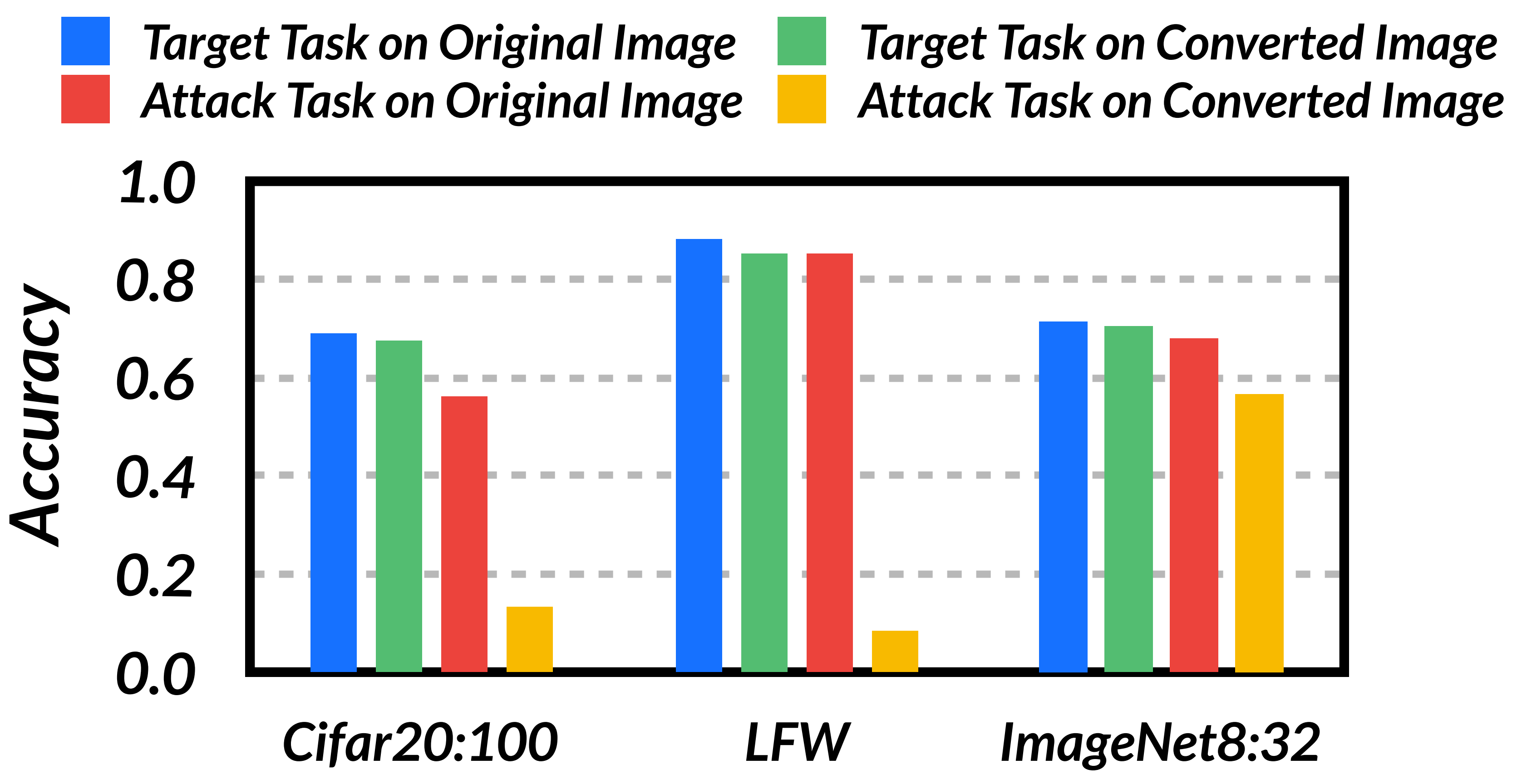
*The adversary can use SOTA DL model to label the entropy-reduced outputs of DAPter.*



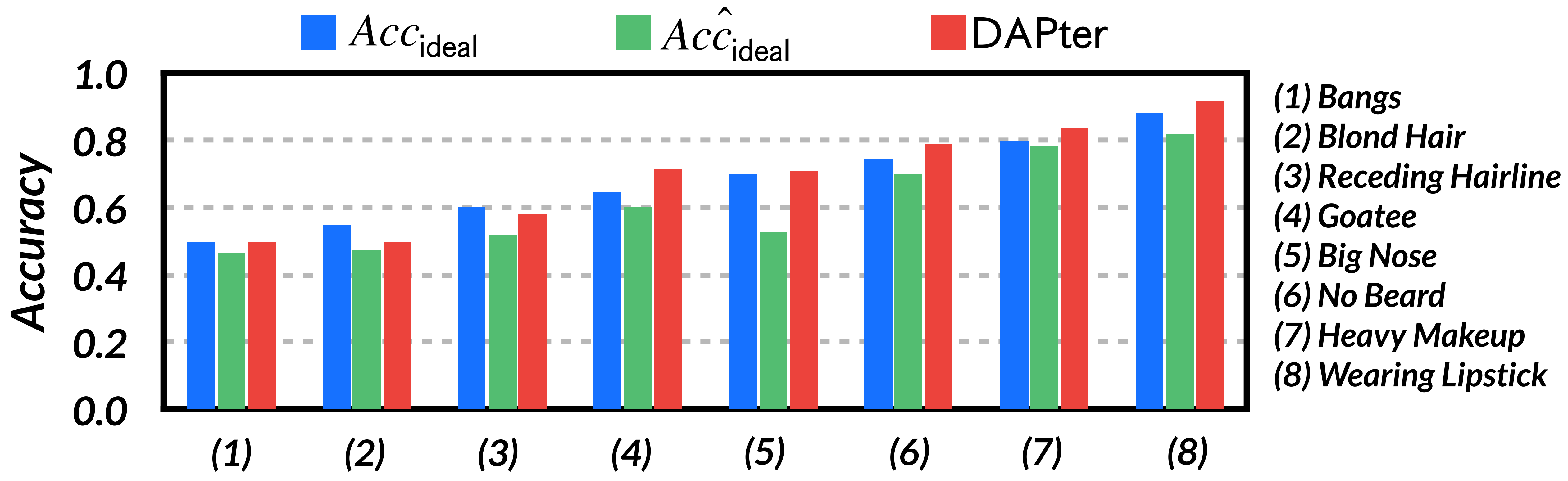
# ***Security*** - ***Auto Recognition Attack***

*The adversary can use SOTA DL model to label the entropy-reduced outputs of DAPter.*



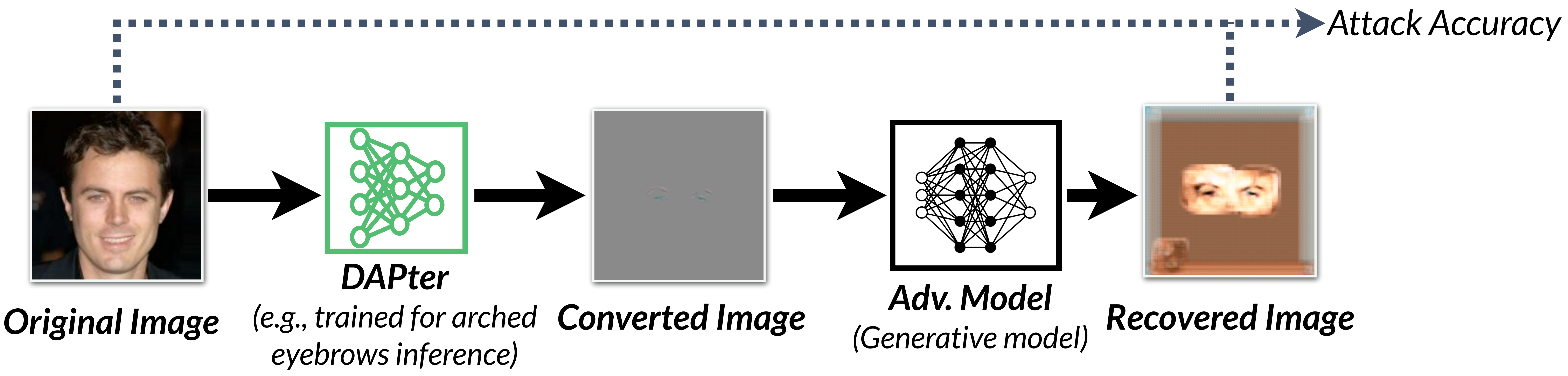

***Case 1:*** *Attack tasks have no correlation with the targeted task.*




***Case 2:*** *Attack tasks have correlations with the targeted task.*

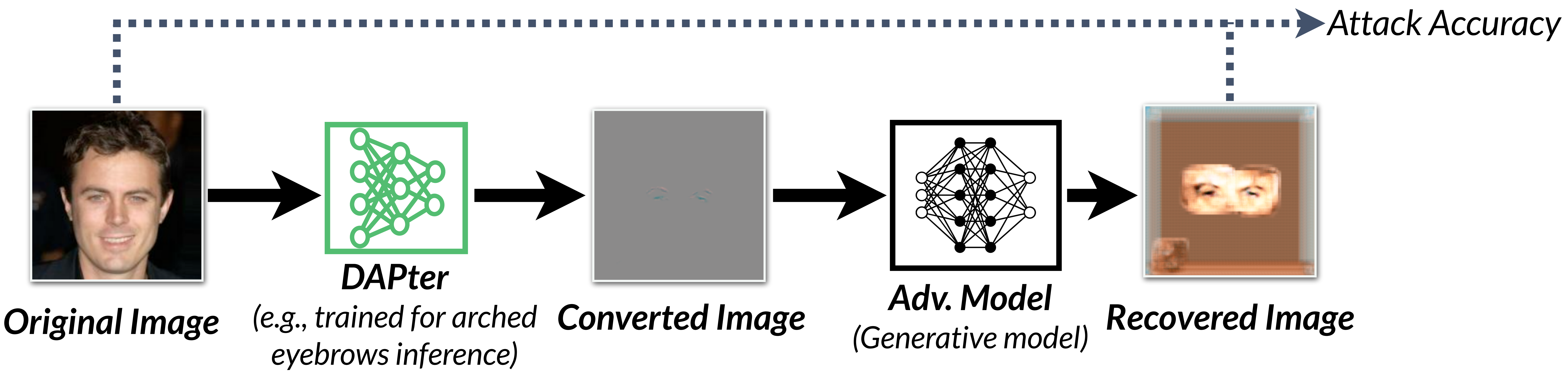
# Security - *Image Reconstruction Attack*

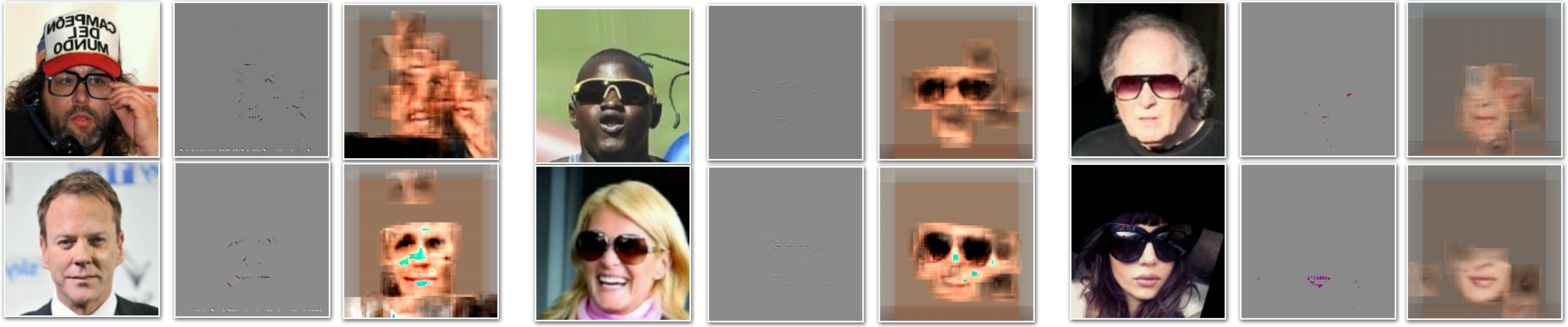
*The adversary can use SOTA DL model to reconstruct the origianl image from the protected one.*



# **Security** - *Image Reconstruction Attack*

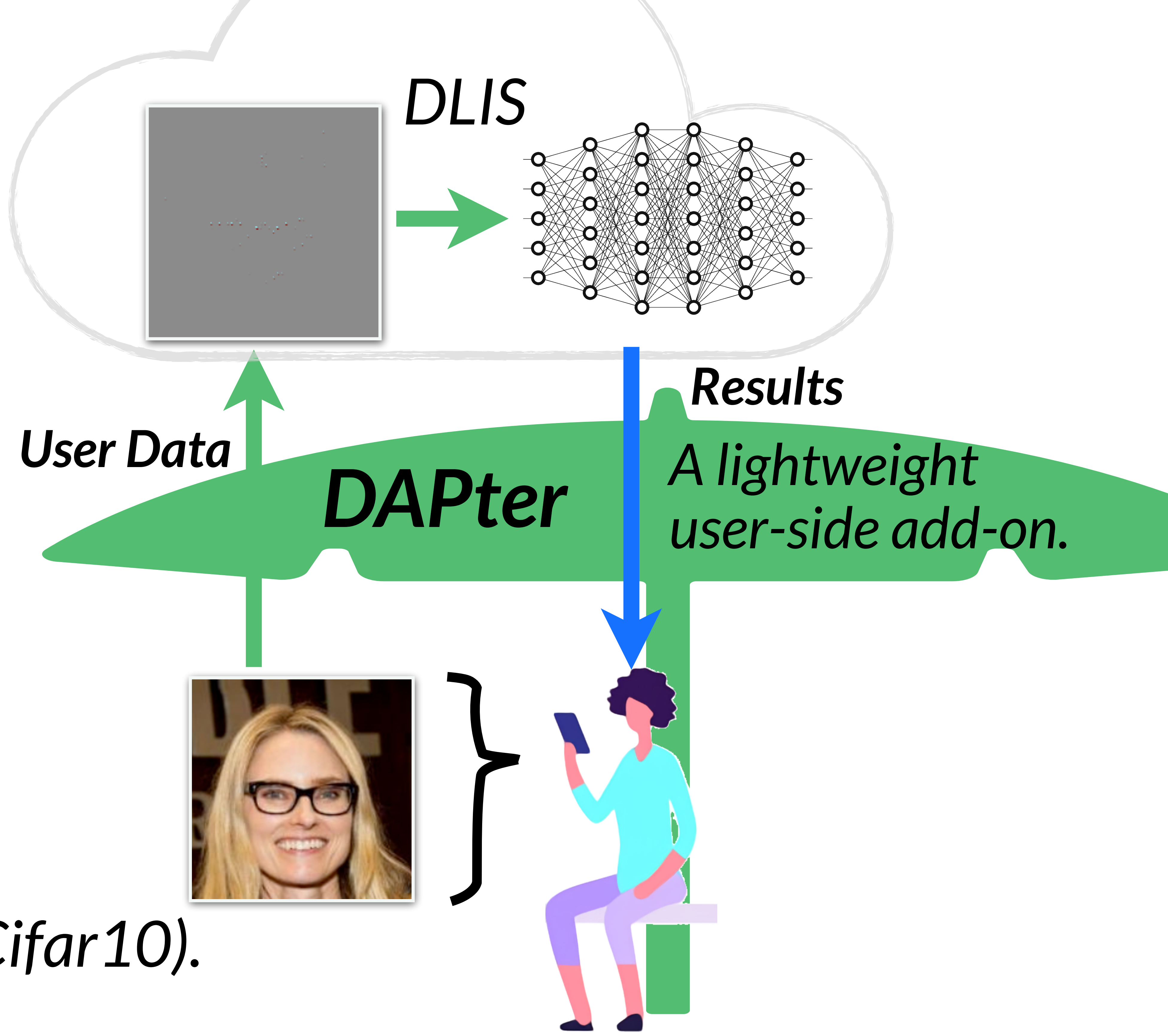
*The adversary can use SOTA DL model to reconstruct the origianl image from the protected one.*



*(a) Chubby Inference Task* *(b) Wearing Glasses Inference Task* *(c) Wearing Lipstick Inference Task*

# *Usability Evaluatuion*




## Backend Throughput:

- *Compare to TEE-based solution:* ***2.5x~50x**,*
- *Compare to FHE-based solution:* ***1000x**.*

## Bandwidth Usage:

- ***2.1x~41x** better (measured with LFW, ImageNet, CelebA, Cifar10).*

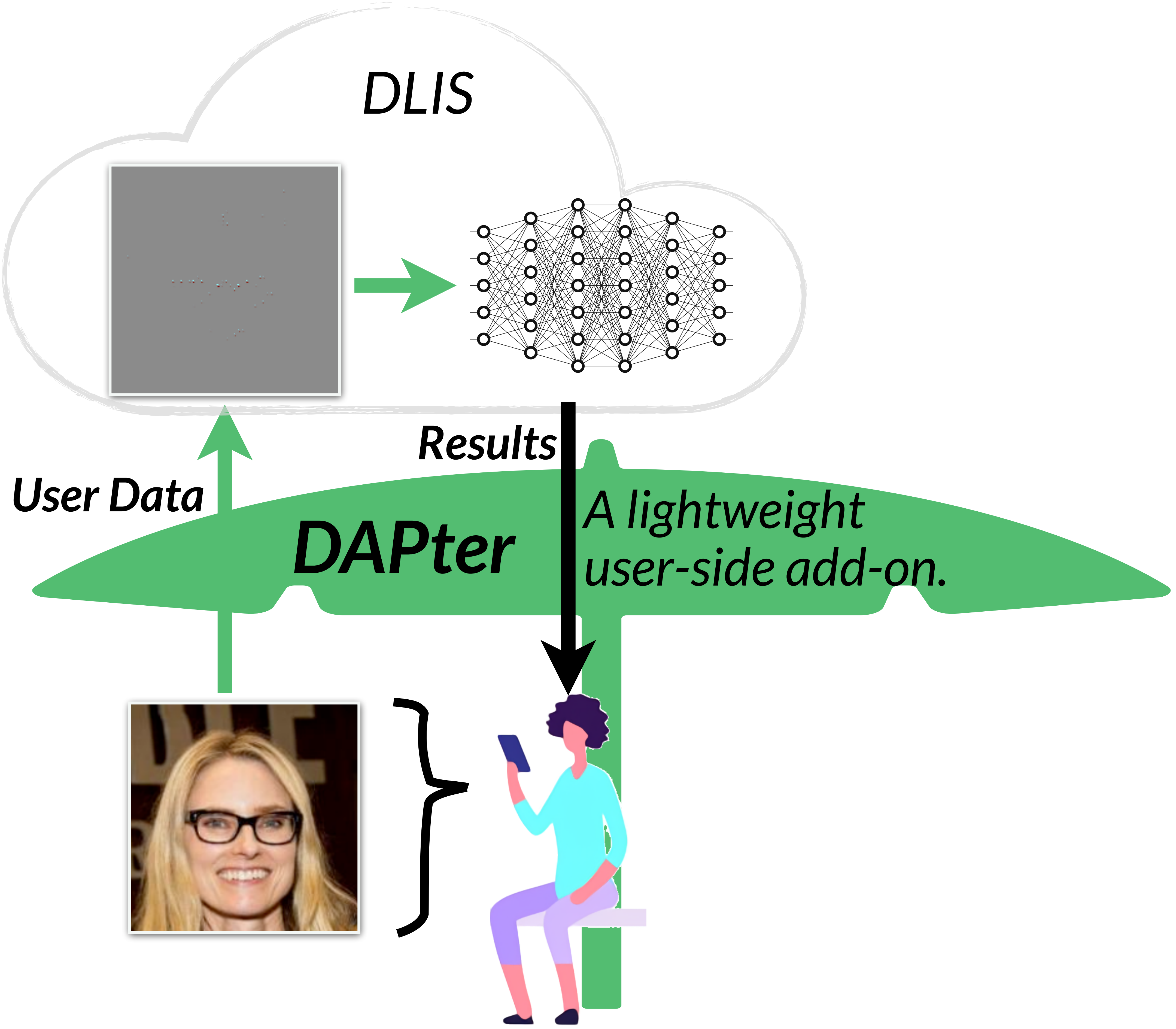
## Latency Overhead:

- ***109ms** (Snapdragon 855 Plus), **292ms** (Kirin 960), and **309ms** (Helio X30).*

***No DLIS backend change is needed!!!***

# Take away

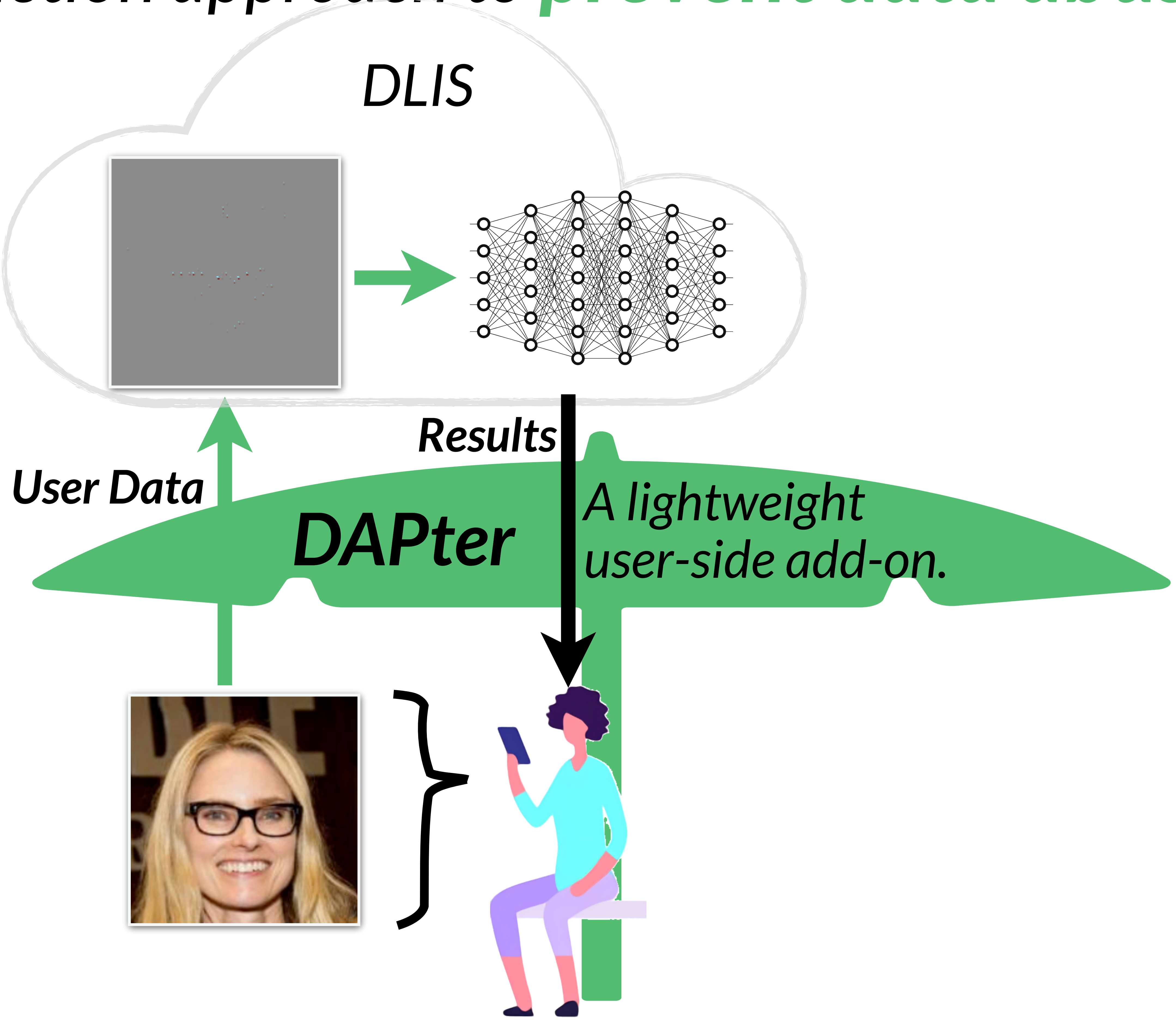
*First investigate the data abuse issue in the scenario of DLIS.*



# Take away

***First*** *investigate the data abuse issue in the scenario of DLIS.*

*A user-side entropy reduction approach to* ***prevent data abuse*** *in DLIS context.*

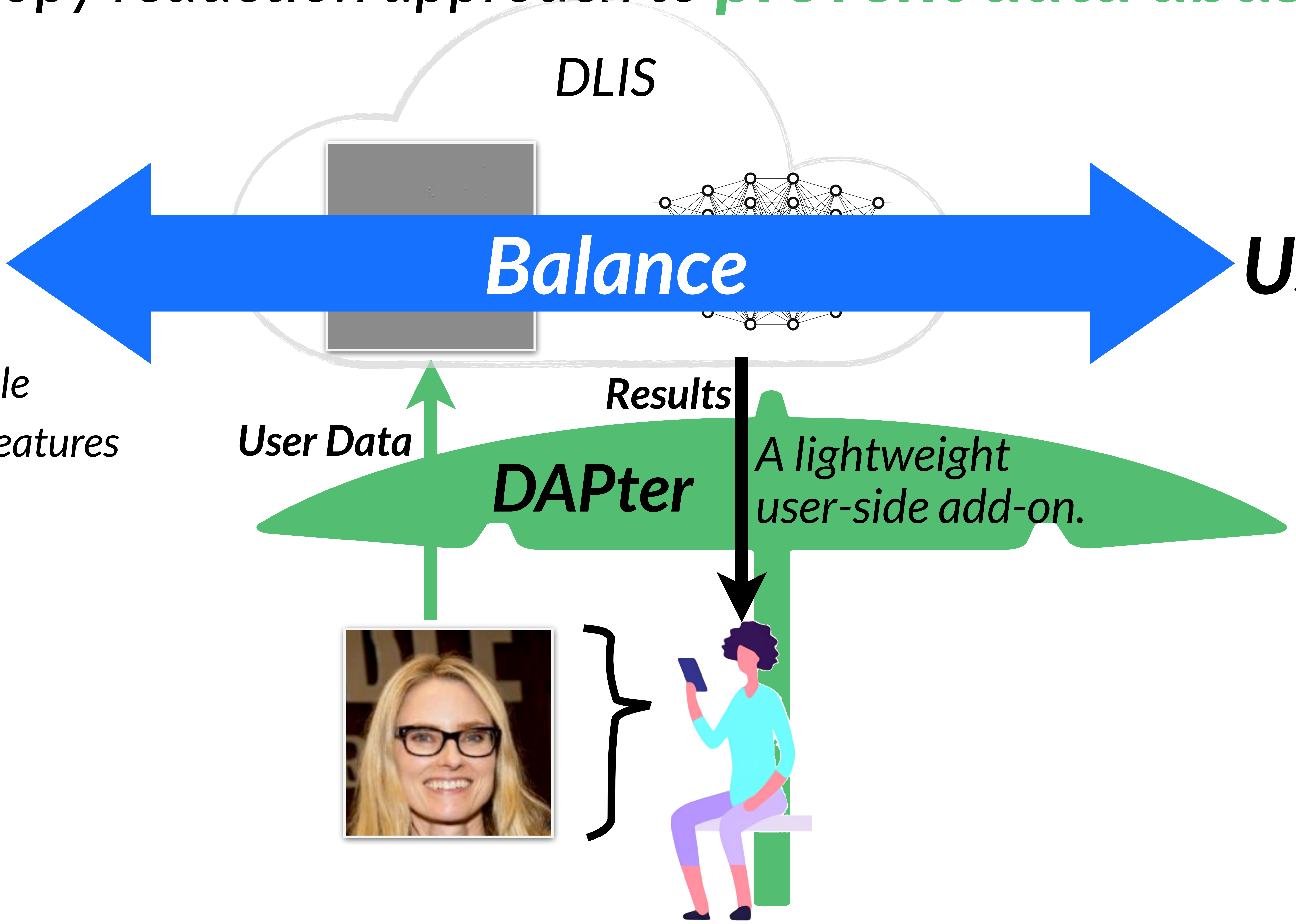


# Take away

*First investigate the data abuse issue in the scenario of DLIS.*

*A user-side entropy reduction approach to **prevent data abuse** in DLIS context.*

**Security**




**Usability**

*S1. Not visually recognizable*
*S2. Only retain necessary features*
*S3. Can't be reversed*

*U1. Maintain Accuracy*
*U2. No Changes*
*U3. Efficient*

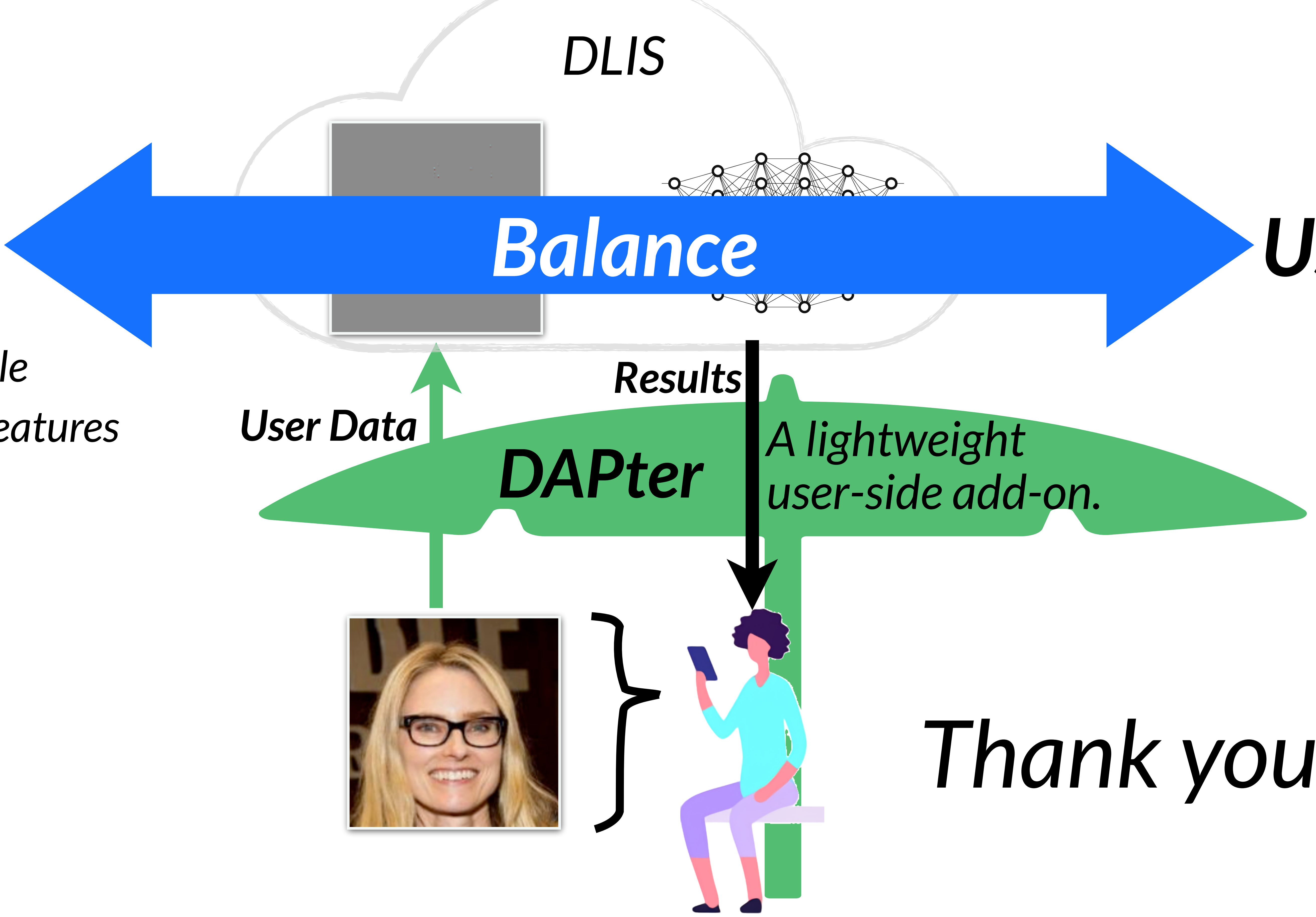
# Take away

***First*** *investigate the data abuse issue in the scenario of DLIS.*

*A user-side entropy reduction approach to* ***prevent data abuse*** *in DLIS context.*

**Security**

- ***S1.*** *Not visually recognizable*
- ***S2.*** *Only retain necessary features*
- ***S3.*** *Can't be reversed*




**Usability**

- ***U1.*** *Maintain Accuracy*
- ***U2.*** *No Changes*
- ***U3.*** *Efficient*

*Thank you for attention!*