MÁSTER EN INTELIGENCIA ARTIFICIAL

DESARROLLO DE UN SISTEMA DE AUTENTICACIÓN BASADO EN LA BIOMETRÍA DEL COMPORTAMIENTO EN DISPOSITIVOS MÓVILES

Curso 2025/26 -

Trabajo dirigido por:
ROBERTO ALCARAZ MACHADO

DAVID JIMÉNEZ CASTRO
DNI: 76052663-N
e-mail: djimenezc1@student.universidadviu.com

Índice general

1. Introducción **1**

1.1. El Fundamento de la Biometría Conductual 1

1.2. Innovación Tecnológica y Arquitectura 2

2. Objetivos **3**

2.1. Objetivo General 3

2.2. Objetivos Específicos 3

2.3. Alcance y Limitaciones 5

3. Estado del Arte **7**

3.1. IJCB 2022 MobileB2C: Competición de Autenticación Basada en Comportamiento Móvil 7

3.2. BehaveFormer: A Framework with Spatio-Temporal Dual Attention Transformers for IMU enhanced Keystroke Dynamics . 9

3.3. Type2Branch: Keystroke Biometrics based on a Dual-branch Architecture with Attention Mechanisms and Set2set Loss . . 11

3.3.1. Evolución de las Arquitecturas de Aprendizaje Profundo 11

3.3.2. Funciones de Pérdida y Verificación 11

3.3.3. Métricas de Rendimiento en Autenticación Pasiva . . . 12

3.3.4. Importancia para la Autenticación Pasiva 12

3.4. Estado del Arte: Autenticación Continua y Ensamblados de Redes Profundas 12

3.4.1. Evolución de las Técnicas de Clasificación 13

3.4.2. Extracción de Características y Representación de Datos 13

3.4.3. Desafíos en la Autenticación Pasiva 14

4. Marco Teórico **17**

4.1. Sistemas de autenticación biométrica 17

4.2. Métodos de aprendizaje automático aplicados 19

4.2.1. Redes de Memoria a Largo Plazo 20

4.2.2. Ventajas frente a RNN's clásicas 21

4.2.3. Transformers y Mecanismos de Atención 22

4.3. Bases de datos y benchmarks del dominio 24

4.4. Privacidad y Consideraciones Éticas 24

4.4.1. Marco Normativo 24

4.4.2. Consideraciones Éticas del Sistema 26

5. Desarrollo del Proyecto **29**

5.1. Descripción del Problema 29

5.2. Metodología 30

5.2.1. Preprocesamiento y Extracción de Características . . . 30

5.2.2. Enfoque 1: Arquitectura basada en Transformer 31

5.2.3. Enfoque 2: Ensemble de Modelos Especializados por Modalidad 32

5.2.4. Protocolo de Evaluación 32

5.3. Resultados Preliminares 33

5.3.1. Baseline de Referencia 33

5.3.2. Estimaciones Esperadas del Sistema Propuesto 33

Bibliografía 35

Capítulo 1

Introducción

En el panorama actual de la ciberseguridad móvil, la dependencia de métodos de autenticación estática —como contraseñas, PINs o patrones de dibujo— presenta vulnerabilidades críticas frente a ataques de ingeniería social e interceptación. Si bien la biometría fisiológica (huella dactilar o reconocimiento facial) ha reforzado el perímetro de seguridad, estos métodos actúan como guardianes de "un solo paso" que no protegen la sesión una vez iniciada.

El presente proyecto propone el desarrollo de un **Sistema de Autenticación Basado en la Biometría del Comportamiento**, un enfoque que permite la verificación de identidad de forma **pasiva, continua y transparente** para el usuario.

1.1. El Fundamento de la Biometría Conductual

A diferencia de los rasgos físicos, la biometría conductual se centra en **cómo** interactúa el usuario con su dispositivo[cite: 26]. Este sistema aprovecha la singularidad de los patrones neurofisiológicos y motores, analizando

variables clave capturadas de manera transparente:

- **Dinámica de Tecleo (*Keystroke Dynamics*):** Estudio de los tiempos de presión (*Dwell Time*), las latencias entre teclas (*Flight Time*) y los intervalos entre pulsaciones (*Inter-key Interval*).
- **Interacción Táctil:** Patrones de desplazamiento (*scrolling*), toques en pantalla (*tapping*) y gestos en la interfaz.
- **Sensores de Movimiento (*IMU*):** Uso de acelerómetros, giroscopios y magnetómetros para capturar la micro-gestualidad y el manejo físico del terminal durante la interacción.

1.2. Innovación Tecnológica y Arquitectura

Para superar las limitaciones de los modelos clásicos (como SVM o KNN), que presentan dificultades de escalabilidad en entornos masivos [cite: 86, 87], este sistema se fundamenta en arquitecturas de **Aprendizaje Profundo (*Deep Learning*)**:

- **Modelos Híbridos y Atención:** Integración de redes neuronales convolucionales (CNN) para la extracción de características locales y redes recurrentes (LSTM) para dependencias temporales globales.
- **Mecanismos de Atención Dual:** Empleo de arquitecturas basadas en *transformers* (como *STDAT*) que aplican atención tanto temporal como de canal para capturar patrones característicos.
- **Fusión Multimodal:** Combinación de la dinámica de tecleo con datos de sensores inerciales para incrementar la robustez del sistema frente a cambios de contexto.
- **Funciones de Pérdida Avanzadas:** Implementación de técnicas como *Triplet Loss* y *Set2set Loss*, permitiendo comparar flujos continuos de datos contra el perfil del usuario con mínima fricción.

Capítulo 2

Objetivos

El presente trabajo tiene como propósito fundamental el diseño, desarrollo y evaluación de un sistema de autenticación continua y pasiva basado en la biometría del comportamiento en dispositivos móviles. A continuación se detallan los objetivos generales y específicos que guiarán la investigación.

2.1. Objetivo General

Desarrollar un sistema de autenticación basado en biometría conductual para dispositivos móviles que sea capaz de verificar la identidad del usuario de forma continua y transparente, sin requerir interacción explícita, mediante el uso de arquitecturas de aprendizaje profundo y fusión multimodal de señales.

2.2. Objetivos Específicos

Para alcanzar el objetivo general, se plantean los siguientes objetivos específicos:

1. **Diseñar una arquitectura de aprendizaje profundo multimodal** capaz de procesar de forma conjunta señales heterogéneas —dinámica

de tecleo (*keystroke dynamics*) y datos de sensores inerciales (IMU)—para generar representaciones de identidad robustas frente a variaciones de contexto y dispositivo.

2. **Implementar y evaluar mecanismos de atención dual** (temporal y de canal) inspirados en arquitecturas tipo Transformer, con el fin de ponderar automáticamente los instantes y características más discriminativos del flujo de interacción del usuario.

3. **Explorar funciones de pérdida orientadas a la verificación**, en particular *Triplet Loss* y *Set2set Loss*, que permitan optimizar el sistema para la comparación de perfiles continuos de usuario en lugar de clasificaciones estáticas.

4. **Evaluar el rendimiento del sistema sobre benchmarks públicos reconocidos** en el dominio, tales como BehavePassDB, HuMIdb o Aalto DB, utilizando métricas estándar como el *Equal Error Rate* (EER) y el *Area Under the Curve* (AUC), con el objetivo de facilitar la comparación con trabajos previos.

5. **Analizar el impacto de la longitud de secuencia** en la precisión de autenticación, buscando alcanzar una detección fiable con ráfagas cortas de actividad (en torno a 50 caracteres), lo que se traduce directamente en una reducción del *Time-to-Detection* (TTD) ante posibles impostores.

6. **Garantizar el cumplimiento de los principios de privacidad y ética** aplicables al tratamiento de datos biométricos conductuales, siguiendo las directrices del RGPD y el principio de *Privacy by Design*, con especial atención a la minimización de datos y la no recuperabilidad del contenido textual a partir de las métricas capturadas.

2.3. Alcance y Limitaciones

El sistema se circunscribe al ámbito de los dispositivos móviles con pantalla táctil, siendo el escenario principal de uso la escritura de texto libre (*free-text*). Quedan fuera del alcance de este trabajo la autenticación basada en biometría fisiológica (huella dactilar, reconocimiento facial) y el análisis de señales de voz o vídeo u otros sistemas de autenticación.

Capítulo 3

Estado del Arte

El estado del arte en la biometría del comportamiento en dispositivos móviles abarca una amplia gama de técnicas y metodologías que buscan identificar y autenticar a los usuarios basándose en sus patrones de interacción con el dispositivo. A continuación, se presentan algunas de las investigaciones y desarrollos más relevantes en este campo.

3.1. IJCB 2022 MobileB2C: Competición de Autenticación Basada en Comportamiento Móvil

El trabajo de Stragapede et al. 2022*c* presenta una **evaluación comparativa** de sistemas de autenticación móvil basados en **biometría conductual**, capturada de manera transparente mientras el usuario interactúa con su dispositivo.

Base de Datos Utilizada

El estudio emplea la base pública **BehavePassDB**, recopilada en condiciones reales, que incluye:

- Dinámica de tecleo (*keystroke*)
- Lectura de texto (*text reading*)
- Deslizamiento de galería (*gallery swiping*)
- Toques en pantalla (*tapping*)
- Sensores como acelerómetro, giroscopio, magnetómetro, entre otros

Para el *benchmarking*, además de la identeficación del usuario legítimo, se consideran dos tipos de ataques de impostores:

- **Impostores aleatorios**: otro usuario con un dispositivo distinto.
- **Impostores hábiles**: individuos que intentan imitar al usuario legítimo.

Conclusiones

Los resultados de la competición **MobileB2C** muestran que la autenticación basada en comportamiento es **viable**, aunque sigue siendo un desafío complejo debido a la variabilidad del entorno y la dificultad de modelar múltiples modalidades de interacción.

Como contribución adicional, el estudio consolida **MobileB2C como una competición continua**, proporcionando una base de datos abierta y un protocolo estándar que facilita nuevas investigaciones en autenticación conductual bajo condiciones realistas.

Cuadro 3.1: Resultados durante la fase de evaluación (AUC [%])

#	Team	Mixed	Random	Skilled
Task 1: Keystroke				
1	NUS-UoA-UoM	**66.37**	**64.77**	**67.91**
2	HCI Essen	51.12	53.02	51.23
3	HBKU CS Lab	51.25	49.38	53.13
4	JAIRG	45.57	52.29	39.89
Task 2: Text Reading				
1	HCI Essen	**57.63**	**61.27**	**53.98**
2	NUS-UoA-UoM	54.89	58.49	51.29
3	JAIRG	50.63	50.00	41.25
4	HBKU CS Lab	48.27	59.42	37.13
Task 3: Gallery Swiping				
1	HBKU CS Lab	**61.54**	**67.35**	**55.73**
2	JAIRG	55.94	61.95	50.62
3	NUS-UoA-UoM	55.66	55.54	55.77
4	HCI Essen	54.72	57.30	51.17
Task 4: Tapping				
1	HBKU CS Lab	**59.58**	**57.22**	**61.94**
2	NUS-UoA-UoM	52.39	54.72	50.06
3	JAIRG	46.25	48.75	43.75
4	HCI Essen	43.89	40.16	47.62

3.2. BehaveFormer: A Framework with Spatio-Temporal Dual Attention Transformers for IMU enhanced Keystroke Dynamics

El trabajo de Senerath et al. 2023*a* propone *BehaveFormer*, un sistema de autenticación continua en dispositivos móviles basado en la biometría del

comportamiento. El enfoque combina dinámicas de tecleo con datos procedentes de sensores inerciales (IMU), presentes de forma estándar en la mayoría de teléfonos inteligentes. El núcleo del modelo es el *Spatio-Temporal Dual Attention Transformer* (STDAT), una arquitectura basada en *transformers* que emplea mecanismos de atención tanto temporal como de canal para capturar patrones característicos en el comportamiento del usuario.

Para la modalidad de tecleo, se utilizan secuencias de di-gramas y trigramas enriquecidas con tiempos de pulsación (*hold*), transiciones entre eventos y latencias entre teclas. En el caso de la IMU, se extraen derivadas de primer y segundo orden sobre los ejes tridimensionales y se aplica la transformada rápida de Fourier (FFT), obteniendo un vector descriptivo de 36 características por instante.

Cada modalidad es procesada por un STDAT independiente; posteriormente ambas representaciones se fusionan mediante concatenación para generar una incrustación final del usuario. El entrenamiento se realiza mediante *triplet loss*, favoreciendo que muestras del mismo usuario queden próximas en el espacio de representación y que las de distintos usuarios estén separadas.

El método se evalúa sobre tres conjuntos de datos ampliamente utilizados en autenticación continua: Aalto DB, HMOG DB y HuMIdb. Los resultados muestran mejoras significativas frente al estado del arte previo, alcanzando por ejemplo un EER del 1.80 % utilizando solo tecleo en Aalto DB, y un EER del 2.95 % al combinar tecleo e IMU en HuMIdb. Estas cifras confirman la eficacia de la fusión multimodal y del mecanismo de atención dual.

En conjunto, BehaveFormer demuestra que la combinación de información de tecleo y sensores inerciales, unida a arquitecturas basadas en transformers, constituye una vía sólida para sistemas de autenticación pasiva y continua, incrementando la seguridad sin exigir interacción explícita del usuario.

3.3. Type2Branch: Keystroke Biometrics based on a Dual-branch Architecture with Attention Mechanisms and Set2set Loss

En esta sección se analiza el artículo de González et al. 2025*a* que introduce *Type2Branch*, un sistema de autenticación basado en la dinámica de tecleo que se destaca por su capacidad de generalización en escenarios reales, superando las limitaciones de modelos anteriores como *TypeNet*.

3.3.1. Evolución de las Arquitecturas de Aprendizaje Profundo

Históricamente, los sistemas de *Keystroke Dynamics* se basaban en redes recurrentes simples para modelar secuencias temporales. Sin embargo, las limitaciones de estas para capturar patrones locales llevaron al desarrollo de enfoques híbridos:

- **Arquitectura Dual-Branch:** El modelo propone una rama de Redes Neuronales Convolucionales (CNN) para la extracción de características locales y una rama de Redes Recurrentes (RNN/LSTM) para las dependencias temporales globales.
- **Mecanismos de Atención:** A diferencia de los modelos precedentes como *TypeNet*, se introducen capas de atención que permiten al sistema ponderar eventos de pulsación específicos que poseen mayor carga discriminativa para la identidad del usuario.

3.3.2. Funciones de Pérdida y Verificación

Uno de los mayores avances del artículo es la transición de funciones de pérdida punto a punto hacia enfoques de conjuntos:

- **Set2set Loss:** Frente al tradicional *Triplet Loss*, la función *Set2set* permite comparar un conjunto de muestras de enrolamiento contra una muestra de consulta. Esto es crítico para la **autenticación pasiva**, donde la decisión se basa en el flujo continuo de datos y no en una única entrada estática.

3.3.3. Métricas de Rendimiento en Autenticación Pasiva

El rendimiento reportado establece un nuevo estándar para escenarios de texto libre (*free-text*) y dispositivos heterogéneos:

Cuadro 3.2: Rendimiento de Type2Branch en entornos reales.

Escenario	**Usuarios**	**EER (%)**	**Longitud de Secuencia**
Desktop (Escritorio)	15,000	0,77 %	50 caracteres
Mobile (Táctil)	5,000	1,03 %	50 caracteres

3.3.4. Importancia para la Autenticación Pasiva

El modelo demuestra que es posible alcanzar una alta tasa de precisión con ráfagas cortas de actividad (50 caracteres). En el contexto de la seguridad continua, esto reduce el *Time-to-Detection* (TTD) de un posible impostor, permitiendo una revocación de acceso casi instantánea sin intervención del usuario.

3.4. Estado del Arte: Autenticación Continua y Ensamblados de Redes Profundas

El estudio de la biometría de teclado (*Keystroke Dynamics*) ha transitado desde el análisis de textos fijos hacia la autenticación pasiva y continua en entornos de texto libre. El trabajo de Acien and et al. 2022 contextualiza este avance mediante la comparación de técnicas tradicionales frente a

arquitecturas de aprendizaje profundo.

3.4.1. Evolución de las Técnicas de Clasificación

La literatura previa se divide fundamentalmente en dos vertientes metodológicas:

- **Aproximaciones Clásicas:** Se basan en algoritmos de aprendizaje automático supervisado como *K-Nearest Neighbors* (KNN), *Naive Bayes* y *Support Vector Machines* (SVM). Aunque eficaces en datasets pequeños, presentan limitaciones de escalabilidad y precisión en escenarios de autenticación continua donde el flujo de datos es masivo y desestructurado.
- **Aprendizaje Profundo (Deep Learning):** El estado del arte reciente destaca el uso de Redes Neuronales Convolucionales (CNN) para extraer características espaciales de las pulsaciones y Redes Neuronales Recurrentes (RNN), específicamente LSTM, para capturar la dependencia temporal. Aversano et al. introducen el concepto de *Ensemble Learning*, combinando múltiples clasificadores base para reducir la varianza y mejorar la robustez del sistema.

3.4.2. Extracción de Características y Representación de Datos

El consenso en las investigaciones actuales identifica tres métricas temporales críticas para la biometría conductual:

1. **Dwell Time (DT):** El tiempo que una tecla permanece presionada.
2. **Flight Time (FT):** El intervalo entre la liberación de una tecla y la pulsación de la siguiente.
3. **Inter-key Interval (IKI):** El tiempo transcurrido entre dos pulsaciones consecutivas.

El artículo subraya que la integración de estos rasgos en arquitecturas profundas permite omitir la ingeniería de características manual, ya que la red aprende representaciones jerárquicas de los patrones de tecleo.

3.4.3. Desafíos en la Autenticación Pasiva

A pesar de los avances, el estado del arte identifica brechas significativas que este estudio busca solventar:

- **Fragmentación de Datos:** La mayoría de los estudios previos utilizan datasets pequeños (menos de 100 usuarios). El artículo destaca la necesidad de *benchmarks* masivos, proponiendo un dataset integrado de más de 160,000 usuarios.
- **Generalización:** La dificultad de los modelos para mantener la precisión cuando el usuario cambia de contexto de escritura o de dispositivo.
- **Toma de Decisión Continua:** La transición de una autenticación de un solo paso (login) a una monitorización constante sin fricción para el usuario.

Estado del Arte: Biometría Conductual en Dispositivos Móviles

El artículo de Stragapede et al. 2022*a* sitúa su investigación en el contexto de la autenticación pasiva mediante biometría conductual en smartphones, analizando tanto la interacción táctil como los sensores de movimiento (background sensors).

Clasificación de Técnicas Anteriores

Los autores clasifican los trabajos previos principalmente en base a las modalidades biométricas y las arquitecturas de procesamiento:

- **Dinámica de Tecleo (Keystroke Dynamics):** Se divide en escenarios de *texto fijo* (contraseñas) y *texto libre*. Se mencionan aproximaciones basadas en redes neuronales recurrentes, específicamente Long Short-Term Memory (LSTM) para autenticación a gran escala.
- **Gestos Táctiles (Touch Gestures):** Técnicas que analizan el desplazamiento (scrolling), toques (tapping) y dibujo de patrones.
- **Sensores de Movimiento (Background Sensors):** Uso de acelerómetros, giroscopios y magnetómetros para capturar patrones físicos del usuario durante la interacción.
- **Arquitecturas de Aprendizaje:** El artículo distingue entre modelos estadísticos tradicionales y enfoques modernos de Deep Learning (como CNNs para gestos y LSTMs para secuencias temporales).

Limitaciones Identificadas

El estudio subraya varias deficiencias en la literatura actual:

1. **Fragmentación de Modalidades:** La mayoría de los estudios se centran en una sola modalidad (solo tecleo o solo acelerómetro), perdiendo la sinergia de la fusión multimodal.
2. **Escalabilidad y Realismo:** Muchos trabajos utilizan bases de datos pequeñas o recolectadas en entornos de laboratorio controlados que no reflejan el uso cotidiano (ïn-the-wild").
3. **Ventanas de Tiempo Elevadas:** Algunos sistemas requieren periodos de observación demasiado largos para alcanzar precisiones aceptables, lo que reduce la eficacia de la autenticación pasiva inmediata.

Problema Específico a Resolver

El vacío que este artículo intenta llenar es la **evaluación exhaustiva de la fusión de múltiples biometrías conductuales** (tecleo, scroll, dibujo, sensores de fondo) bajo un protocolo común y utilizando una base de datos

pública de gran escala (*HuMIdb*). El objetivo es determinar cuál es la combinación mínima de sensores y tiempo de análisis necesaria para lograr una autenticación robusta y transparente.

Capítulo 4

Marco Teórico

4.1. Sistemas de autenticación biométrica

La biometría se ha consolidado como el pilar fundamental de la identidad digital moderna, desplazando a los sistemas basados en posesión y conocimiento. Este marco teórico analiza la biometría no solo como una herramienta técnica, sino como un proceso científico de reconocimiento de individuos basado en sus características biológicas y conductuales.

La biometría se basa en la premisa de que ciertos rasgos físicos o de comportamiento son únicos, estables y medibles. Para que una característica humana sea considerada un indicador biométrico eficaz, debe cumplir con siete criterios fundamentales definidos por Jain et al. 2008 (Sección 1.6 *biometric characteristics*):

1. **Universalidad**: Cada persona debe poseer el rasgo.
2. **Unicidad**: Dos personas no deben tener el mismo rasgo (distinguibilidad).
3. **Permanencia**: El rasgo debe ser *invariante* en el tiempo.
4. **Colectabilidad**: El rasgo debe poder medirse cuantitativamente.

5. **Rendimiento**: Precisión y velocidad en el reconocimiento.
6. **Aceptabilidad**: Grado en que la población está dispuesta a utilizar el sistema.
7. **Resistencia al fraude**: Dificultad para engañar al sistema mediante artefactos.

Los sistemas biométricos se basan en un flujo de trabajo estándar dividido en dos fases críticas:

1. **Alistamiento o Registro**: En esta fase un sistema de sensores captura la huella biométrica del individuo. Posteriormente se elimina el ruido y se extraen las características de interés de la huella. Finalmente se genera una representación matemática de la huella. Es importante remarcar que no se guarda la imagen original de la huella sino su representación matemática.
2. **Reconocimiento**: En esta fase se distinguen dos procesos:
   - **Verificación**: Se realiza la comparación *uno a uno* de la huella (*¿Es, realmente, la persona que dice ser?*).
   - **Identificación**: Se realiza la comparación *uno a varios* de la huella (*¿Quién es esta persona?*).

(Incluir un diagrama de los dos sistemas quedaría bien)

Modalidades de biometría conductual

En la literatura científica es común ver que la biometría se clasifica en tres grupos:

1. **Biometría Fisiológica**: Se centra en la medida directa de diferentes partes del cuerpo como: la huella dactilar, el reconocimiento facial, ocular, geometría de la mano y su patrón de venas.

2. **Biometría Conductual**: Analiza patrones de comportamientos que pueden ser aprendidos o adquiridos como: la mecanografía, la marcha (forma de caminar) y la firma.

3. **Biometría Química**: Se basa en las características químicas del individuo como el ADN o el olor corporal (compuestos órganicos volátiles emitidos por la piel).

En este trabajo nos centraremos en la biometría conductual o del comportamiento.

Métricas de Rendimiento y Errores

Para medir la eficacia del sistema biométrico usaremos principalmente las siguientes tasas de error fundamentales:

1. **FAR (*False Acceptance Rate*)**: Probabilidad de que el sistema acepte por error a un impostor.

2. **FRR (*False Rejection Rate*)**: Probabilidad de que el sistema rechace erróneamente a un usuario legítimo.

3. **ERR (*Equal Error Rate*)**: El punto donde FAR y FRR son iguales.

4.2. Métodos de aprendizaje automático aplicados

La evolución de la biometría moderna está intrínsecamente ligada al avance del *Aprendizaje Automático* y del *Aprendijaze Profundo*. Estas arquitecturas permiten pasar de una comparación basada en reglas manuales a una extracción de características latentes con dimensionalidad alta.

4.2.1. Redes de Memoria a Largo Plazo

Las redes neuronales LSTM *(Long Short-Term Memory)* son una variante de las Redes Neuronales Recurrentes *(RNN)* diseñadas para evitar el problema del desvanecimiento del gradiente *(vanishing gradient)*, permitiendo aprender dependencias a largo plazo en datos secuenciales Hochreiter and Schmidhuber 1997.

Arquitectura y Mecanismo de Compuertas

A diferencia de una RNN estándar, cuya celda oculta se actualiza completamente en cada paso temporal, la LSTM introduce un estado de celda C_t que actúa como una "memoria" capaz de retener información relevante durante largas secuencias. Este mecanismo se regula mediante tres compuertas diferenciables:

- **Compuerta de olvido** f_t: Decide qué información del estado anterior debe descartarse. Se define como:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{4.1}$$

- **Compuerta de entrada** i_t: Controla qué nueva información se almacena en el estado de celda:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4.2}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{4.3}$$

- **Compuerta de salida** o_t: Determina qué parte del estado de celda se expone como salida:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{4.4}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{4.5}$$

Donde σ es la función sigmoide, W son las matrices de pesos entrenables, b los sesgos, h_{t-1} el estado oculto anterior y x_t la entrada en el instante t. La actualización del estado de celda combina el olvido selectivo del pasado con la incorporación de nueva información:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{4.6}$$

donde $\odot$ denota el producto elemento a elemento *(Hadamard)*.

4.2.2. Ventajas frente a RNN's clásicas

El problema del desvanecimiento del gradiente en RNN estándar surge cuando, durante la retropropagación a través del tiempo *(BPTT)*, el gradiente se multiplica repetidamente por valores menores que uno, haciendo que la señal de error se extinga antes de alcanzar pasos temporales lejanos. Las LSTM mitigan este efecto gracias al estado de celda, que permite que el gradiente fluya sin atenuación a través de largos intervalos temporales siempre que la compuerta de olvido permanezca abierta Goodfellow et al. 2016.

Aplicación a la Biometría Conductual

En el contexto de la autenticación basada en biometría del comportamiento, las LSTM resultan especialmente adecuadas por la naturaleza inherentemente secuencial de las señales capturadas. Para la **dinámica de tecleo**, cada pulsación genera un vector de características temporales —Dwell Time (DT), Flight Time (FT) e Inter-key Interval (IKI)— que forman una secuencia variable en longitud. La LSTM es capaz de modelar las dependencias entre pulsaciones no adyacentes, capturando el ritmo global de escritura del usuario, algo imposible de representar con modelos estáticos como SVM o KNN Acien et al. 2022*a*.

Del mismo modo, para señales de **sensores inerciales** (acelerómetro y giroscopio), la LSTM aprende patrones de micro-gestualidad que se manifies-

tan en ventanas temporales de cientos de milisegundos, extrayendo representaciones del comportamiento físico del usuario al sostener e interactuar con el dispositivo.

Trabajos como TypeNet Acien et al. 2022*a* y BehaveFormer Senerath et al. 2023*b* han demostrado que las LSTM alcanzan resultados competitivos en autenticación continua de texto libre, siendo posteriormente complementadas con mecanismos de atención y arquitecturas basadas en Transformers para capturar dependencias tanto locales como globales en la secuencia de interacción.

4.2.3. Transformers y Mecanismos de Atención

Originalmente propuestos por Vaswani et al. para el procesamiento del lenguaje natural, los Transformers han revolucionado múltiples campos de la inteligencia artificial gracias a su capacidad para modelar relaciones globales entre todos los elementos de una secuencia de forma paralela, superando las limitaciones de las arquitecturas recurrentes en cuanto a eficiencia computacional y captura de dependencias a larga distancia Vaswani et al. 2017.

El Mecanismo de Atención

El núcleo del Transformer es el mecanismo de *Scaled Dot-Product Attention*, que opera sobre tres matrices: consultas *(Queries, Q)*, claves *(Keys, K)* y valores *(Values, V)*, todos ellos proyecciones lineales de la entrada. La atención se calcula como:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4.7}$$

donde d_k es la dimensión de las claves, y el factor $\frac{1}{\sqrt{d_k}}$ evita que el producto escalar crezca excesivamente para dimensiones altas, estabilizando el gradiente durante el entrenamiento Vaswani et al. 2017.

En la práctica se emplea *Multi-Head Attention*, que aplica h cabezas de atención en paralelo sobre distintas proyecciones del espacio de representación, permitiendo al modelo atender simultáneamente a diferentes aspectos de la secuencia:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{4.8}$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{4.9}$$

Atención Dual Espacio-Temporal en Biometría Conductual

En el contexto de la autenticación basada en biometría del comportamiento, el mecanismo de atención adquiere una relevancia especial, ya que no todas las pulsaciones o instantes temporales contribuyen por igual a la identidad del usuario. Arquitecturas como BehaveFormer Senerath et al. 2023*b* introducen el *Spatio-Temporal Dual Attention Transformer (STDAT)*, que aplica dos dimensiones de atención de forma simultánea:

- **Atención Temporal**: Pondera la relevancia de cada instante en la secuencia de pulsaciones, identificando qué momentos del flujo de tecleo son más discriminativos para verificar la identidad.
- **Atención de Canal**: Opera sobre las características extraídas de los sensores inerciales (IMU), determinando qué ejes del acelerómetro o giroscopio aportan mayor información identitaria en cada contexto de uso.

Fusión Multimodal mediante Atención

Una de las ventajas más relevantes de los Transformers para este trabajo es su capacidad natural para la **fusión multimodal**. Al tratar cada modalidad —dinámica de tecleo, gestos táctiles y señales IMU— como una secuencia de tokens, el mecanismo de atención aprende automáticamente a qué fuente de información debe dar mayor peso en cada instante, optimizando la decisión de autenticación sin necesidad de definir manualmente reglas de fusión Senerath et al. 2023*b*. Esto contrasta con los enfoques clásicos de

fusión a nivel de puntuación, donde la combinación de modalidades se realiza mediante reglas heurísticas fijas.

Type2Branch González et al. 2025*b* extiende este principio incorporando capas de atención sobre una arquitectura dual CNN-LSTM, permitiendo al modelo ponderar eventos de pulsación específicos con mayor carga discriminativa, lo que se traduce en mejoras significativas del Equal Error Rate (EER) frente a modelos precedentes como TypeNet.

4.3. Bases de datos y benchmarks del dominio

Aquí incluiré lo que utilice para comparar más adelante...

4.4. Privacidad y Consideraciones Éticas

El desarrollo de sistemas de autenticación basados en biometría conductual implica la captura, procesamiento y almacenamiento de datos de carácter personal de especial sensibilidad. En consecuencia, cualquier implementación debe enmarcarse en el conjunto normativo vigente en materia de protección de datos, tanto a nivel europeo como nacional.

4.4.1. Marco Normativo

Reglamento General de Protección de Datos (RGPD)

El Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, conocido como RGPD Parlamento Europeo and Consejo de la Unión Europea 2016, establece el marco general de protección de datos personales en la Unión Europea. En su artículo 9, el RGPD clasifica los **datos biométricos** como una categoría especial de datos personales, cuyo tratamiento queda prohibido con carácter general salvo que concurra alguna de las excepciones tasadas, entre las que destacan:

- El **consentimiento explícito** del interesado para uno o varios fines específicos.
- La necesidad del tratamiento por razones de **interés público esencial**.
- Fines de **investigación científica**, siempre que se apliquen las garantías adecuadas.

Adicionalmente, el RGPD consagra una serie de principios de obligado cumplimiento para cualquier sistema que trate datos biométricos conductuales:

- **Minimización de datos**: únicamente deben recogerse los datos estrictamente necesarios para el fin declarado. En el contexto de este sistema, esto implica limitar la captura a las métricas temporales de tecleo (DT, FT, IKI) y señales IMU imprescindibles, evitando el almacenamiento del contenido textual introducido por el usuario.
- **Limitación de la finalidad**: los datos recogidos no podrán utilizarse para fines distintos a la autenticación del usuario para el que fueron enrolados.
- **Exactitud y actualización**: los perfiles biométricos conductuales deben mantenerse actualizados, dado que los patrones de comportamiento del usuario pueden evolucionar con el tiempo debido a factores como el estado emocional, la fatiga o el cambio de dispositivo.
- **Limitación del plazo de conservación**: los datos biométricos no deben conservarse más tiempo del necesario para cumplir la finalidad del tratamiento.
- **Integridad y confidencialidad**: el sistema debe garantizar la seguridad de los datos mediante cifrado y controles de acceso adecuados.

Ley Orgánica de Protección de Datos y Garantía de los Derechos Digitales (LOPDGDD)

En el ámbito nacional español, la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales Jefatura del Estado 2018 complementa y adapta el RGPD al ordenamiento

jurídico español. En materia de datos biométricos, la LOPDGDD refuerza las obligaciones del responsable del tratamiento y establece garantías adicionales, entre las que destacan:

- La obligatoriedad de realizar una **Evaluación de Impacto relativa a la Protección de Datos (EIPD)** antes de implementar cualquier sistema que trate datos biométricos a gran escala, tal y como exige el artículo 35 del RGPD.
- El reconocimiento explícito del **derecho a la portabilidad** de los perfiles biométricos, permitiendo al usuario solicitar la transferencia o eliminación de su perfil conductual en cualquier momento.
- La figura del **Delegado de Protección de Datos (DPD)**, cuya designación puede ser obligatoria cuando el tratamiento de datos biométricos se realice a gran escala.

4.4.2. Consideraciones Éticas del Sistema

Más allá del cumplimiento normativo estricto, el diseño de un sistema de autenticación conductual plantea una serie de consideraciones éticas que deben abordarse desde las fases tempranas del desarrollo.

Transparencia y Consentimiento Informado

Dado que la biometría conductual opera de forma pasiva y transparente para el usuario, existe el riesgo de que este no sea plenamente consciente de los datos que se están recopilando. Por ello, es imprescindible garantizar que el usuario reciba una información clara, comprensible y accesible sobre el funcionamiento del sistema antes de otorgar su consentimiento, evitando el uso de cláusulas genéricas o formulaciones técnicas ininteligibles.

Privacidad desde el Diseño

El principio de *Privacy by Design* Cavoukian 2009, incorporado al artículo 25 del RGPD como *protección de datos desde el diseño y por defecto*, exige que las medidas de privacidad se integren en la arquitectura del sistema desde su concepción, y no como un añadido posterior. En la práctica, esto se traduce en decisiones de diseño como:

- Almacenar únicamente representaciones vectoriales cifradas del perfil conductual, nunca las secuencias de pulsaciones en bruto.
- Aplicar técnicas de **anonimización** o **seudonimización** que impidan vincular el perfil biométrico con la identidad real del usuario en caso de brecha de seguridad.
- Garantizar que el modelo entrenado no permita reconstruir el contenido textual original a partir de las métricas temporales capturadas.

Sesgos y Equidad del Sistema

Los modelos de aprendizaje profundo pueden presentar sesgos en su rendimiento en función de características del usuario como la edad, condición motora o nivel de familiaridad con el dispositivo. Un sistema de autenticación conductual que presente tasas de error significativamente distintas entre grupos de usuarios podría generar situaciones de discriminación inadvertida. Por ello, es necesario evaluar el rendimiento del sistema de forma estratificada y adoptar medidas correctoras cuando se detecten disparidades.

Revocabilidad del Perfil Biométrico

A diferencia de la biometría fisiológica, los patrones conductuales son, en cierta medida, revocables: si el perfil de un usuario se ve comprometido, es posible reentrenar el modelo con nuevas muestras. No obstante, esto implica mantener protocolos claros de gestión del ciclo de vida del perfil biométrico, incluyendo procedimientos de re-enrolamiento y eliminación segura de perfiles obsoletos.

Capítulo 5

Desarrollo del Proyecto

5.1. Descripción del Problema

El problema central de este trabajo es la autenticación continua y pasiva de usuarios en dispositivos móviles a partir de su biometría conductual. A diferencia de los mecanismos de autenticación estática —contraseñas, PINs o biometría fisiológica— que únicamente verifican la identidad en el momento del acceso, el objetivo es mantener una verificación activa durante toda la sesión sin que el usuario deba realizar ninguna acción adicional. Este enfoque plantea un problema de **verificación biométrica de una clase** (*one-class* o *open-set*): dado un perfil conductual enrolado para un usuario legítimo, el sistema debe decidir, a partir de una ventana corta de interacción, si el individuo que está operando el dispositivo corresponde a dicho perfil o se trata de un impostor. La dificultad inherente reside en que durante la fase de entrenamiento no se dispone de muestras etiquetadas de todos los posibles impostores, lo que impide formular el problema como una clasificación cerrada convencional. Las modalidades conductuales seleccionadas para este trabajo son tres, elegidas por su complementariedad y por la riqueza de información identitaria que aportan:

- **Dinámica de tecleo (*keystroke dynamics*)**: secuencias de tiempos

de pulsación (*Dwell Time*, DT), tiempos de vuelo entre teclas (*Flight Time*, FT) e intervalos entre pulsaciones consecutivas (*Inter-key Interval*, IKI). Constituye la modalidad más estudiada en la literatura y ofrece alta discriminabilidad incluso con ráfagas cortas de texto.

- **Dinámica de firma (*signature dynamics*)**: patrones de presión, velocidad y aceleración durante el trazo de la firma en pantalla táctil. Aporta una señal de alta varianza inter-usuario y baja varianza intra-usuario cuando el individuo firma de forma natural.
- **Gestos de desplazamiento (*scroll dynamics*)**: características extraídas de los movimientos de deslizamiento sobre la pantalla, incluyendo velocidad, aceleración, longitud del trazo y ángulo de desplazamiento. Su inclusión está condicionada al análisis del beneficio marginal que aporta en la fusión multimodal.

El principal desafío técnico es la **heterogeneidad temporal** de las señales: la dinámica de tecleo es discreta y de longitud variable, la firma es una señal continua de corta duración, y el scroll produce eventos esporádicos dependientes del contexto de uso. El sistema debe ser capaz de integrar estas tres fuentes de forma coherente y tomar decisiones de autenticación con latencia mínima, idealmente tras menos de 50 eventos de interacción.

5.2. Metodología

Para abordar el problema descrito, se propone una metodología experimental comparativa basada en dos enfoques arquitectónicos complementarios, cuyo rendimiento se evaluará bajo un protocolo común.

5.2.1. Preprocesamiento y Extracción de Características

Previo al modelado, cada modalidad requiere un pipeline de preprocesamiento específico:

- **Tecleo**: a partir de los eventos de pulsación se extraen las tripletas (DT_i, FT_i, IKI_i) para cada tecla i de la secuencia. Las secuencias se segmentan en ventanas deslizantes de longitud fija N (objetivo: $N = 50$ eventos), con solapamiento configurable para la autenticación continua. Los valores atípicos se filtran mediante umbrales basados en percentiles y las secuencias se normalizan con estadísticos calculados por usuario en la fase de enrolamiento.
- **Firma**: se extraen series temporales de posición (x_t, y_t), presión p_t y sus derivadas de primer y segundo orden. La señal se remuestrea a una frecuencia fija para garantizar la homogeneidad de la representación de entrada.
- **Scroll**: se caracterizan cada gesto de deslizamiento mediante un vector de rasgos agregados: velocidad media, velocidad máxima, aceleración media, longitud euclidiana del trazo, duración y desviación angular respecto al eje vertical.

5.2.2. Enfoque 1: Arquitectura basada en Transformer

El primer enfoque propone una arquitectura *end-to-end* inspirada en BehaveFormer Senerath et al. 2023*b* y Type2Branch González et al. 2025*b*, que procesa las tres modalidades de forma conjunta mediante mecanismos de atención. Cada secuencia de interacción se trata como una serie de *tokens* de entrada. Un codificador Transformer independiente por modalidad —con capas de *Multi-Head Self-Attention* y redes *feed-forward* posicionales— genera una representación de contexto para cada fuente de señal. Las representaciones resultantes se fusionan mediante un módulo de **atención cruzada entre modalidades** (*cross-modal attention*), que aprende a ponderar dinámicamente la contribución de cada fuente en función del contexto de uso. La función de pérdida empleada para el entrenamiento es *Set2set Loss* González et al. 2025*b*, que compara un conjunto de muestras de enrolamiento $\mathcal{S} = \{s_1, \ldots, s_k\}$ contra una muestra de consulta q, permitiendo una estimación más robusta de la similitud que el clásico *Triplet Loss* puntual:

$$\mathcal{L}_{\text{set2set}} = \text{máx}\left(0,\ m + d\big(\bar{e}_{\mathcal{S}},\, q^{+}\big) - d\big(\bar{e}_{\mathcal{S}},\, q^{-}\big)\right) \tag{5.1}$$

donde $\bar{e}_{\mathcal{S}}$ es la representación agregada del conjunto de enrolamiento, q^{+} una muestra del usuario legítimo, q^{-} una muestra de impostor y m el margen de separación.

5.2.3. Enfoque 2: Ensemble de Modelos Especializados por Modalidad

El segundo enfoque adopta una estrategia modular: se entrena un modelo especializado e independiente para cada modalidad conductual, y sus puntuaciones de similitud se combinan mediante una estrategia de fusión a nivel de *score*.

- **Modelo de tecleo**: red LSTM bidireccional seguida de una capa de proyección métrica, entrenada con *Triplet Loss*, siguiendo la línea de TypeNet Acien et al. 2022*b*.
- **Modelo de firma**: red CNN-LSTM que extrae características locales de los segmentos del trazo (CNN) y modela la dinámica temporal global (LSTM).
- **Modelo de scroll**: clasificador ligero basado en un perceptrón multicapa (MLP) o una red CNN-1D sobre los vectores de rasgos agregados por gesto.

La fusión de puntuaciones se explorará mediante tres estrategias: combinación lineal con pesos fijos, combinación lineal con pesos aprendidos, y un meta-clasificador entrenado sobre las puntuaciones de los modelos base.

5.2.4. Protocolo de Evaluación

Ambos enfoques se evaluarán bajo un protocolo común de verificación, siguiendo las directrices establecidas en las competiciones MobileB2C Stra-

gapede et al. 2022*d*:

- **Partición de datos**: división en conjuntos de enrolamiento, desarrollo y evaluación, sin solapamiento de usuarios entre particiones.
- **Escenarios de impostor**: se considerarán impostores aleatorios (usuarios distintos sin conocimiento del objetivo) e impostores hábiles (usuarios que intentan imitar al legítimo), cuando el dataset seleccionado lo permita.
- **Métricas principales**: *Equal Error Rate* (EER) como métrica primaria de comparación, complementada con AUC, FAR@1 %FRR y el *Time-to-Detection* (TTD) estimado.

5.3. Resultados Preliminares

En el momento de redacción de este documento, el desarrollo del sistema se encuentra en fase de diseño arquitectónico y planificación experimental. Por tanto, los resultados que se presentan a continuación son estimaciones de referencia (*baseline*) derivadas de la replicación parcial de trabajos del estado del arte sobre datasets públicos, con el objetivo de establecer un punto de comparación para las arquitecturas propuestas.

5.3.1. Baseline de Referencia

A continuación se muestra la tabla que recoge los valores de EER reportados por los sistemas más relevantes del estado del arte sobre las modalidades y datasets que se emplearán en este trabajo, y que constituyen el umbral de rendimiento que el sistema propuesto deberá superar.

5.3.2. Estimaciones Esperadas del Sistema Propuesto

A modo orientativo, y en base a las mejoras arquitectónicas propuestas respecto a los sistemas de referencia, se anticipan los rangos de rendimiento

Cuadro 5.1: Resultados de referencia del estado del arte (EER %).

Sistema	**Modalidad**	**Dataset**	**EER (%)**
TypeNet Acien et al. 2022*b*	Tecleo	Aalto DB	2.20
Type2Branch González et al. 2025*b*	Tecleo	Mobile	1.03
BehaveFormer Senerath et al. 2023*b*	Tecleo + IMU	HuMIdb	2.95
MobileB2C top-1 Stragapede et al. 2022*d*	Keystroke	BehavePassDB	AUC: 66.37

recogidos en la siguiente tabla:

Cuadro 5.2: Estimación de rendimiento esperado del sistema propuesto (EER %). **[PENDIENTE DE ACTUALIZAR]**

Enfoque	**Modalidades**	**EER esperado (%)**	**Observacion**
Transformer unificado	Tecleo + Firma	∼1.5 – 2.5	Fusión cross-mo
Transformer unificado	Tecleo + Firma + Scroll	∼1.2 – 2.0	Mejora con scr
Ensemble modular	Tecleo + Firma	∼1.8 – 3.0	Fusión de scor
Ensemble modular	Tecleo + Firma + Scroll	∼1.5 – 2.5	Meta-clasificad

Estas estimaciones se fundamentan en la hipótesis de que la fusión multimodal de tecleo y firma aportará una mejora consistente respecto al uso de tecleo en solitario, tal y como sugieren los trabajos previos en fusión de modalidades conductuales Stragapede et al. 2022*b*. La incorporación del scroll se plantea como un experimento ablativo para cuantificar su contribución marginal al rendimiento global del sistema.

Bibliografía

Acien, A. y et al.: 2022, *Expert Systems with Applications*

Acien, A., Morales, A., Vera-Rodriguez, R., Tolosana, R., y Fierrez, J.: 2022a, *IEEE Transactions on Biometrics, Behavior, and Identity Science* **4(1)**, 57

Acien, A., Morales, A., Vera-Rodriguez, R., Tolosana, R., y Fierrez, J.: 2022b, *IEEE Transactions on Biometrics, Behavior, and Identity Science* **4(1)**, 57

Cavoukian, A.: 2009

González, N., Stragapede, G., Vera-Rodriguez, R., y Tolosana, R.: 2025a, *Type2Branch: Keystroke Biometrics based on a Dual-branch Architecture with Attention Mechanisms and Set2set Loss*

González, N., Stragapede, G., Vera-Rodriguez, R., y Tolosana, R.: 2025b, *arXiv preprint*

Goodfellow, I., Bengio, Y., y Courville, A.: 2016, *Deep Learning*, MIT Press, Cambridge, MA

Hochreiter, S. y Schmidhuber, J.: 1997, *Neural Computation* **9(8)**, 1735

Jain, A. K., Flynn, P. J., y Ross, A. A. (eds.): 2008, *Handbook of Biometrics*, Springer Science+Business Media, LLC, New York, NY, USA

Jefatura del Estado: 2018, *Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales*

Parlamento Europeo y Consejo de la Unión Europea: 2016, *Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos*

Senerath, D., Tharinda, S., Vishwajith, M., Rasnayaka, S., Wickramanayake, S., y Meedeniya, D.: 2023a, *BehaveFormer: A Framework with Spatio-Temporal Dual Attention Transformers for IMU enhanced Keystroke Dynamics*

Senerath, D., Tharinda, S., Vishwajith, M., Rasnayaka, S., Wickramanayake, S., y Meedeniya, D.: 2023b, in *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*

Stragapede, G., Vera-Rodriguez, R., Tolosana, R., Morales, A., Acien, A., y Le Lan, G.: 2022a, *Pattern Recognition Letters* **157**, 35

Stragapede, G., Vera-Rodriguez, R., Tolosana, R., Morales, A., Acien, A., y Le Lan, G.: 2022b, *Pattern Recognition Letters* **157**, 35

Stragapede, G., Vera-Rodriguez, R., Tolosana, R., Morales, A., Fierrez, J., Ortega-Garcia, J., Rasnayaka, S., Seneviratne, S., Dissanayake, V., Liebers, J., Islam, A., Belhaouari, S. B., Ahmad, S., y Jabin, S.: 2022c, *IJCB 2022 Mobile Behavioral Biometrics Competition (MobileB2C)*

Stragapede, G., Vera-Rodriguez, R., Tolosana, R., Morales, A., Fierrez, J., Ortega-Garcia, J., Rasnayaka, S., Seneviratne, S., Dissanayake, V., Liebers, J., Islam, A., Belhaouari, S. B., Ahmad, S., y Jabin, S.: 2022d, in *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., y Polosukhin, I.: 2017, in *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30